MONICA ALBU

Construirea și utilizarea testelor psihologice



PREFAŢĂ

Rolul important al matematicii în toate domeniile vieţii şi ale ştiinţei a fost recunoscut de foarte multă vreme. Leonardo da Vinci susţinea că "nu există nici o certitudine în ştiinţele în care nu este posibilă aplicarea vreuneia din ştiinţele matematice", iar A. Einstein afirma că "matematica le oferă ştiinţelor naturii o doză de siguranţă, pe care acestea nu ar putea să o obţină altfel".

Statistica, la rândul ei, are o aplicabilitate generală în domeniul ştiinţific. "În toate cercetările ştiinţifice de bază, statisticienii sunt cei care elaborează planurile de experienţă sau de observaţii şi tot la ei se recurge pentru a analiza rezultatele, pentru a evalua constatările şi pentru a separa faptele clar demonstrate de cele care mai cer confirmarea", observa Ronald Fischer.

La fel ca şi celelalte discipline ştiinţifice, psihologia foloseşte modele matematice şi statistice la explorarea relaţiilor dintre variabile şi la testarea ipotezelor. Dar, datorită complexităţii şi ambiguităţii unor aspecte ale comportamentului uman, cu ajutorul procedeelor de calcul obişnuite este dificilă realizarea unor inferenţe valide şi precise. Din acest motiv, au fost create modele noi, specifice ştiinţelor sociale, cum sunt teoria răspunsului la itemi, folosită la construirea testelor psihologice, şi teoria variabilelor latente (care cuprinde analiza factorială şi modelele cu ecuaţii structurale), utilă pentru studierea relaţiilor dintre variabilele măsurate prin testele psihologice.

Lucrarea de faţă are ca principal obiectiv familiarizarea cititorului cu metodele matematice şi statistice care se folosesc la construirea testelor psihologice şi la interpretarea scorurilor testelor. Alături de procedeele de calcul uzuale sunt prezentate metode noi, bazate pe teoria aproximării şi pe teoria clasificării. Acestea îmbină ideile teoriei clasice a testelor cu cele ale teoriei răspunsului la itemi.

Căutându-se răspuns la întrebarea "Ce înseamnă un test bun?" s-a realizat o sinteză a ideilor cuprinse în lucrări clasice şi moderne din domeniul teoriei testelor psihologice. Este propus şi un punct de vedere nou asupra teoriei testelor, bazat pe teoria clasificării: testul psihologic este privit ca un instrument de ierarhizare şi de clasare a subiecţilor, iar construirea unui test psihologic este asimilată operaţiei de creare a unui clasament prin agregarea clasamentelor produse de itemi.

Lucrarea oferă, în plus, informaţii despre condiţiile de utilizare a testelor psihologice. Sunt discutate aspecte privind cine, când şi cum poate folosi un test psihologic.

Două anexe conţin definiţiile noţiunilor matematice şi statistice care intervin în metodele de calcul prezentate. Ele sunt însoţite de exemple.

Toţi algoritmii proprii propuşi au fost verificaţi pe loturi de subiecţi. La administrarea testelor în acest scop şi la interpretarea, sub aspect psihologic, a rezultatelor obţinute am primit ajutor din partea domnului prof. dr. Horia Pitariu.

Cartea se adresează în primul rând studenţilor de la facultăţile de psihologie, dar ea le poate fi utilă şi psihologilor care, în activitatea lor curentă, fac uz de teste psihologice.

CUPRINS

CAPITOLUL 1
TEST ŞI TESTARE PSIHOLOGICĂ

1.1. CE ESTE UN TEST PSIHOLOGIC ?

Testul este un instrument al metodei experimentale, folosit cu precădere în investigaţiile cu caracter aplicativ ale psihologiei (testul psihologic), pedagogiei (testul pedagogic, testul şcolar), sociologiei (testul sociometric) şi biologiei (testul fiziologic).

Testul psihologic reprezintă, alături de observaţii, interviuri, scale de evaluare, tehnici proiective etc., o sursă de informare utilizată pentru evaluarea comportamentului uman şi a proceselor mintale.

În privinţa definiţiei testului psihologic, nu s-a ajuns la un consens, fiecare autor oprindu-se asupra altor aspecte sau reţinând alte caracteristici. De exemplu:

- Testul psihologic este o măsurătoare obiectivă şi standardizată a unui eşantion de comportamente (Anastasi, 1976).
- Testul psihologic este o situaţie standardizată care generează un comportament semnificativ (Klausnitzer, 1992).
- Testul psihologic constă dintr-o probă, mai frecvent dintr-o serie de probe, construite în scopul stabilirii prezenţei (sau a absenţei) unui aspect psihic, a particularităţilor de comportare sau a gradului de dezvoltare psihică (Roşca, 1972).
- Testul psihologic este o probă, utilizată mai ales în psihologia diferenţială, care permite descrierea comportamentului unui subiect într-o situaţie definită precis ("instrucţiunile" testului), prin raportare la comportamentul unei grupe de subiecţi fixată, plasată în aceeaşi situaţie. Această descriere se face de obicei sub formă numerică (Grand dictionnaire de la psychologie, 1994).
- Un test psihologic este o procedură standardizată prin care se formează un eşantion de comportamente care va fi descris prin categorii sau scoruri. În plus, multe teste au norme sau standarde care fac posibilă utilizarea rezultatelor la predicţia altor comportamente, mai importante (Gregory, 1992).
- Un test psihologic este definit prin următoarele caracteristici: materiale şi proceduri standardizate; motivaţie optimă; înregistrare imediată; cotare obiectivă; norme potrivite şi validitate verificată (Dahlstrom, 1993).

Fără a încerca să formulăm o definiţie, vom prezenta elementele pe care le considerăm esenţiale pentru înţelegerea noţiunii de "test psihologic".

A. Testele psihologice servesc fie la *măsurarea unor atribute psihice* (ca, de exemplu, atenţia), fie la *predicţia unor comportamente* (ca, de pildă, succesul profesional).

B. Un test psihologic constă din una sau mai multe probe. Fiecare probă determină o reacţie comportamentală a subiectului (cum ar fi, de pildă, reproducerea unui şir de numere) sau îi cere acestuia să facă aprecieri asupra unui comportament al său (de exemplu, să indice frecvenţa întârzierilor la serviciu ori intensitatea efortului depus pentru îndeplinirea la timp a sarcinilor de serviciu). Răspunsurile sunt evaluate, conform unor reguli fixate, obţinându-se una sau mai multe valori numerice, denumite **scoruri**.

Un element al testului psihologic constituit dintr-o situaţie stimul care solicită un răspuns ce poate fi evaluat separat de restul testului se numeşte **item**. Un item poate consta, de exemplu, dintr-o întrebare însoţită de mai multe răspunsuri posibile, dintr-un desen care trebuie completat sau dintr-o listă de termeni care trebuie ordonaţi după un criteriu specificat. Răspunsului dat de subiect la un item i se atribuie un scor.

Un set de itemi care se referă la acelaşi aspect (un atribut psihic pe care testul îl măsoară sau un comportament pe care testul îl prezice) formează o **scală**. Pentru fiecare scală se calculează un scor, pe baza scorurilor itemilor componenţi, conform unei reguli stabilite de autorul testului. De exemplu, scorul unei scale dintr-un test de cunoştinţe poate fi egal cu suma scorurilor itemilor rezolvaţi corect sau cu scorul celui mai dificil item rezolvat corect sau cu scorul celui mai uşor item rezolvat greşit.

Nu există restricţii privind numărul de itemi care formează o scală sau numărul de scale care intră în componenţa unui test psihologic. Multe chestionare de personalitate cuprind zeci de scale, fiecare conţinând mai mulţi itemi. În schimb, alte teste (cum este, de pildă, Testul de capacităţi organizatorice al lui A. Huth) sunt alcătuite dintr-o singură scală, cu un singur item.

C. Pentru a nu necesita un timp de administrare prea lung, nici un test psihologic nu evaluează toate comportamentele care ar putea fi utilizate la definirea sau la măsurarea unui anumit atribut, ci doar un *eşantion de comportamente* din această mulţime, ales astfel încât să fie *reprezentativ* pentru toate comportamentele mulţimii respective care ar putea fi observate în afara situaţiei de testare.

De exemplu, dacă s-ar dori să se evalueze deprinderea copiilor din clasa întâi de a face calcule aritmetice, procedeul corect de lucru ar fi să se urmărească toate calculele efectuate de fiecare subiect: la şcoală, la rezolvarea temelor, la joacă, la cumpărături etc. Evident, acest lucru este imposibil. S-ar putea, însă, construi un test alegând exerciţii care să solicite efectuarea operaţiilor aritmetice cu numere formate din una sau două cifre. Pentru ca

eşantionul să fie reprezentativ, exerciţiile cuprinse în test trebuie să conţină toate cele patru operaţii aritmetice.

Nu este necesar ca itemii unui test construit în scopul predicţiei unui comportament să descrie exact comportamentul respectiv. De altfel, ar fi imposibil ca un test utilizat la selecţia profesională, deci înainte de ocuparea unui loc de muncă sau a unui loc la un curs de instruire pentru o anumită profesie, să se refere la comportamentele care sunt specifice profesiei respective. De exemplu, nu li se poate cere persoanelor care se prezintă la selecţia pentru şcoala de şoferi să conducă o maşină pe şosea. *Testul psihologic măsoară comportamente prin care se pot face inferenţe asupra altor comportamente.*

D. În etapa de construcţie a testului psihologic se stabilesc şi regulile care trebuiesc respectate la utilizarea acestuia, pentru a se asigura o funcţionare corectă a testului ca instrument de măsură. Aceste reguli, care sunt prezentate în manualul testului, se referă la:

- conţinutul testului (care sunt stimulii prezentaţi pentru a provoca reacţiile comportamentale evaluate de test);
- condiţiile de administrare (instrucţiunile date subiectului în legătură cu sarcina ce trebuie executată);
- tehnica de evaluare a răspunsurilor (modul de cotare a reacţiilor, cotare care să nu permită decât o intervenţie redusă a subiectivităţii examinatorului şi, în consecinţă, să realizeze un acord cât mai mare între examinatori);
- modul de interpretare a rezultatelor obţinute de un subiect.

O probă sau o serie de probe formează un test psihologic numai dacă este însoţită de *regulile de utilizare*, adică dacă este *standardizată.*

Vom explica în continuare importanţa respectării regulilor de utilizare a testului.

- Majoritatea testelor au o structură fixă: tuturor subiecţilor li se administrează aceiaşi itemi, în aceeaşi ordine. Dacă testul conţine atât itemi uşori cât şi dificili, iar ordinea itemilor nu ar fi fixă, atunci comportarea la test a subiecţilor ar putea fi influenţată de poziţia itemilor dificili: prezenţa acestora la început poate fi obositoare şi chiar descurajatoare, repercutându-se negativ asupra modului de răspuns la ceilalţi itemi.

Există şi teste al căror conţinut se stabileşte în timpul administrării lor, în funcţie de răspunsurile date de subiect.

- Condiţiile de administrare a testului afectează comportamentul celor testaţi. O persoană are şanse mai mici să răspundă corect la itemii unui test de atenţie atunci când acesta este administrat într-o încăpere de trecere, decât dacă testarea se face într-o cameră liniştită.

În majoritatea cazurilor modul în care este administrat testul depinde şi de competenţa examinatorului (de gradul de informare al acestuia relativ la test, de experienţa pe care o are în administrarea testelor, de grija pe care o manifestă faţă de examinaţi).

Dacă testul este administrat colectiv şi sunt urmate întocmai instrucţiunile din manualul testului, li se poate asigura tuturor subiecţilor o similaritate a condiţiilor de testare. În schimb, la testarea individuală personalitatea examinatorului are un rol mult mai important, fiind posibil ca aceeaşi persoană, cu ocazia a două testări efectuate de examinatori diferiţi, să aibă comportamente total diferite.

• Cele mai multe teste psihologice au reguli fixe pentru calculul scorurilor, formulate astfel încât scorurile obţinute de subiecţi să nu depindă de persoana care le calculează. În această categorie intră, de exemplu, testele cu alegere multiplă, întrucât pentru ele este stabilit, încă din faza de construcţie a lor, punctajul care se acordă fiecărui răspuns propus, la fiecare item. Există şi teste pentru care scorul reflectă aprecierea subiectivă a celui care îl calculează. Un exemplu îl constituie testele de creativitate care cer să se indice cât mai multe posibilităţi de utilizare pentru un anumit obiect, iar punctajul acordat fiecărui răspuns este stabilit de examinator, în funcţie de cât de neobişnuită i se pare utilizarea propusă.

Pentru a cerceta concordanţa aprecierilor făcute de mai mulţi evaluatori, D. Stark şi R. C. Elliot (Vernon, 1965) au trimis câte o copie a unei lucrări de geometrie tuturor profesorilor de matematică de la 16 şcoli, cerându-le să îi acorde un punctaj în conformitate cu modul lor obişnuit de notare. Rezultatele obţinute au variat între 28 şi 92 de puncte. Autorii experimentului au încercat să explice variabilitatea mare a notelor prin existenţa mai multor aspecte care sunt luate în considerare la stabilirea punctajului: corectitudinea soluţiei, originalitatea sa, claritatea redactării lucrării etc. În absenţa unor instrucţiuni pentru acordarea notei, fiecare examinator atribuie alte ponderi acestor aspecte.

Testele eseu, care solicită formularea unui răspuns mai amplu la o anumită problemă, intră în aceeaşi categorie, a testelor cu cotare subiectivă.

• Scorurile obţinute de o persoană, doar prin valoarea lor, nu aduc nici o informaţie. Chiar ştiind că scorurile unei scale care măsoară memoria pot fi cuprinse între 0 şi 30, este imposibil de apreciat dacă un rezultat egal cu 20 indică o memorie obişnuită sau una foarte bună.

Interpretarea unui scor individual se poate face numai prin compararea sa, fie cu rezultatele obţinute de alte persoane care fac parte din aceeaşi populaţie ca şi subiectul, fie cu anumite praguri, indicate de autorul testului, care delimitează mai multe categorii. De exemplu, în cazul testelor de cunoştinţe utilizate la concursurile de admitere, care furnizează un punctaj cuprins între 1

şi 10, se fixează un singur prag, de obicei egal cu 5, iar subiectul este considerat "admis" dacă scorul său este mai mare decât pragul, şi "respins", în caz contrar.

Uneori, pe baza scorurilor obţinute la scalele unui test, se identifică o categorie în care este potrivit să fie repartizată persoana examinată. Categoria poate reprezenta un tip de temperament, o boală psihică, un program de pregătire profesională etc.

Modul de interpretare a rezultatelor este stabilit de autorul testului şi este prezentat în manualul testului.

E. Se spune că un test este *obiectiv* dacă el permite să se măsoare fără ambiguitate capacităţile unei persoane (Klausnitzer, 1992). Pentru a fi obiectiv, testul trebuie să îndeplinească trei condiţii:

- Rezultatele să nu fie dependente de nivelul de pregătire al subiectului în domeniul testării psihologice, adică persoana examinată să nu poată ghici cum se interpretează răspunsurile la itemi şi, prin urmare, să nu poată denatura rezultatele (să nu poată da acele răspunsuri care generează despre sine impresia pe care o doreşte).
- Fiecare răspuns posibil la un item să fie interpretat la fel (ca fiind corect sau greşit, favorabil sau nefavorabil pentru aspectul măsurat etc.) de mai mulţi experţi care analizează testul. De exemplu, la itemul "Mă simt bine într-o societate veselă" mai mulţi psihologi să fie de acord că răspunsul "Adevărat" corespunde persoanelor extravertite, iar răspunsul "Fals" celor introvertite.
- Scorul obţinut de subiect să nu fie influenţat de personalitatea examinatorului, adică testul să fie cu cotare obiectivă. Această condiţie nu este îndeplinită de multe teste, în special de cele proiective şi de testele de cunoştinţe ai căror itemi sunt cu răspunsuri deschise.

F. O trăsătură comună tuturor testelor psihologice este precizia lor limitată. Sunt extrem de rare cazurile când un test furnizează o măsură exactă a unei variabile care are efecte importante asupra comportamentului uman. Din acest motiv, nici deciziile care se iau pe baza lor nu sunt totdeauna corecte. De exemplu, unii psihologi susţin că, din cauza lipsei de consistenţă a comportamentului de la o situaţie la alta, coeficienţii de corelaţie liniară dintre scorurile unui test de personalitate şi un comportament pe care acesta îl prezice nu depăşesc valoarea de 0.3 sau 0.4 (W. Mischel şi R. E. Nisbett, după Funder, 1983). Dar, nu există alte metode de măsurare, mai precise. Prin urmare, *testele psihologice reprezintă cea mai bună, cea mai corectă şi cea mai precisă tehnologie disponibilă pentru a lua decizii importante asupra indivizilor* (Murphy, Davidshofer, 1991).

1.2. UTILIZĂRI ALE TESTELOR PSIHOLOGICE

Testele psihologice sunt folosite, în principal, pentru *a stabili un diagnostic psihologic*, prin evaluarea comportamentului, a abilităţilor mintale şi a altor caracteristici de personalitate, pentru *a face aprecieri şi predicţii* referitoare la subiecţi şi pentru *a lua decizii* asupra persoanelor. În general, testele predictive şi cele pentru decizie se utilizează în selecţia profesională, iar cele diagnostice servesc la măsurarea unor caracteristici psihice, fie pentru aprecierea efectelor unui program de instruire, fie pentru cunoaşterea individului de către psihologul clinician.

Predicţia se referă la estimarea performanţei viitoare a individului, de exemplu la un anumit loc de muncă.

Într-un sens restrâns, însă, chiar şi diagnosticul situaţiei prezente, de pildă al retardării mintale sau al dezechilibrului emoţional, implică o predicţie asupra a ceea ce va face individul în situaţii diferite de cea actuală. Din acest motiv, se poate afirma că *toate testele psihologice sunt eşantioane de comportament pe baza cărora se pot face predicţii asupra altor comportamente* (Anastasi, 1976).

Pe de altă parte, luarea unei decizii corecte se bazează de obicei pe efectuarea unei prognoze asupra consecinţelor deciziei. De exemplu, într-o selecţie de personal decizia de încadrare pe un post se ia în urma comparării performanţelor trecute şi actuale ale indivizilor cu un set de condiţii, considerate a fi absolut necesare pentru îndeplinirea muncii respective. Scopul deciziei este de a identifica acele persoane care vor realiza cele mai bune performanţe pe postul pe care vor fi încadrate. Variabilele care servesc ca bază pentru decizie şi asupra cărora sunt fixate condiţiile eliminatorii trebuie să aibă valoare predictivă.

În funcţie de scopul urmărit prin administrarea testelor, se pot distinge următoarele categorii de utilizări ale acestora (Gregory, 1992):

A. Clasificarea persoanelor

Pe baza rezultatelor obţinute la test se decide în care clasă, din mai multe existente, să fie repartizat fiecare subiect. În acest mod se procedează atunci când:

- se alege cel mai potrivit program de pregătire şcolară sau profesională, în funcţie de particularităţile psihice şi de cunoştinţele persoanelor;
- la sfârşitul unui program de instruire sau cu ocazia examinării pentru admiterea la o formă de învăţământ, se decide dacă un candidat poate fi declarat "reuşit" sau "nereuşit";

• se identifică persoanele care au caracteristici speciale (de exemplu sunt capabile de performanţe superioare într-un anumit domeniu), folosind un test de selecţie simplu, care se administrează rapid (screening test); pentru persoanele care au realizat scoruri mai mari decât un prag fixat se vor administra apoi alte teste, care realizează măsurători mai precise.

B. Diagnoză psihologică şi alegerea tratamentului

Diagnoza urmăreşte să determine natura, cauza şi intensitatea comportamentului anormal al persoanei şi să identifice categoria de boală în care este cel mai potrivit să fie încadrat subiectul (diagnosticul). Pe baza informaţiilor furnizate de diagnoză se instituie tratamentul pentru remedierea deficienţelor sau pentru îmbunătăţirea performanţelor.

Adesea, testele de personalitate servesc la diagnoza tulburărilor emoţionale.

C. Autocunoaştere

Comparând rezultatele obţinute la testele psihologice cu cele ale altor persoane, subiectul se poate cunoaşte mai bine.

D. Evaluarea eficienţei unui program educaţional sau social

O modalitate de evaluare a eficienţei constă în compararea rezultatelor obţinute la teste psihologice de subiecţii care au fost supuşi influenţei programului, fie cu rezultatele unor persoane care nu au fost incluse în programul respectiv, fie cu scorurile aceloraşi persoane, înainte de acţiunea programului.

E. Cercetări teoretice sau aplicative asupra comportamentului uman

Se folosesc diverse teste psihologice pentru a depista cauza unor comportamente sau condiţiile - de natură psihică - în care un anumit tratament (de exemplu, o metodă de psihoterapie sau un program de instruire) are şanse de succes.

Construirea unui test psihologic impune efectuarea unei cercetări în care se folosesc şi teste psihologice cunoscute. Un acelaşi test psihologic poate fi utilizat în scopuri diferite. De exemplu, un test care permite formularea unui diagnostic psihiatric poate servi şi la autocunoaştere.

1.3. CONDIŢII DE UTILIZARE A TESTELOR PSIHOLOGICE

În mai multe ţări au fost elaborate norme etice şi standarde de utilizare a testelor psihologice şi de practicare a examinărilor psihologice.

Standardele citate mai frecvent în literatura de specialitate şi la care vom face referire în această lucrare sunt cele elaborate în S.U.A., în 1985, de reprezentanţi a trei organizaţii profesionale: *The American Educational Research Association*, *The American Psychological Association* (APA) şi *The National Council on Measurement in Education*. Intitulate *Standards for Educational and Psychological Testing*, ele descriu amănunţit condiţiile care trebuiesc îndeplinite la construirea, evaluarea, administrarea şi interpretarea testelor psihologice. Sunt vizate două aspecte, şi anume:

- caracteristicile psihometrice ale testelor;
- comportamentul profesional al psihologilor, insistându-se asupra ideii că ei poartă responsabilitatea pentru apărarea bunăstării persoanelor testate.

Aceste Standarde au fost completate cu prevederi specifice unor domenii particulare. Aşa sunt, de exemplu:

- *Specialy Guidelines* (APA, 1981), cu îndrumări pentru psihologii care lucrează în clinică, în organizaţii industriale, în şcoli sau în activitatea de consiliere;
- *Principles for the Validation and Use of Personnel Selection Procedures* (Society for Industrial and Organizational Psychology, 1987), privind condiţiile de desfăşurare a selecţiei profesionale;
- *Guidelines for Computer-Based Tests and Interpretations* (APA, 1986), referitoare la folosirea calculatorului în testarea psihologică.

În ceea ce priveşte principiile etice care trebuie să îi călăuzească pe psihologi în activitatea de evaluare, în completarea celor cuprinse în *Standards for Educational and Psychological Testing* vin cele enunţate în *Ethical Principles of Psychologists and Code of Conduct* (APA, 1992). Redăm câteva dintre ideile enunţate aici:

- evaluarea psihologică şi psihodiagnosticul trebuie să se facă numai într-un cadru profesional, de către utilizatori de teste competenţi şi cu experienţă în domeniu, cu ajutorul unor teste potrivite;
- pentru proiectarea şi selectarea tehnicilor adecvate populaţiilor particulare se vor folosi proceduri ştiinţifice;
- interpretarea rezultatelor şi explicarea acestora se vor face cu mare atenţie;
- se va acţiona pentru păstrarea securităţii testelor.

În proiectul *Statutului psihologului din România*, propus de biroul Asociaţiei Psihologilor din România, se precizează că profesia de psiholog "se

exercită numai de către psihologi" (art. 1). Această profesie constă "în observarea, evaluarea, explicarea, predicţia şi modificarea comportamentului uman prin aplicarea principiilor, metodelor şi procedurilor ştiinţifice în scopul prevenirii şi tratamentului tulburărilor psihice şi de comportament, precum şi în scopul dezvoltării relaţiilor interpersonale, adaptării eficiente la viaţă şi muncă, optimizării ergonomice a producţiei, creşterii eficienţei personale şi menţinerii sănătăţii psihice şi fizice (art. 2).

Vom prezenta în continuare câteva aspecte care sunt deosebit de importante atunci când se utilizează teste psihologice (Anastasi,1976; Gregory, 1992; Murphy, Davidshofer, 1991).

A. Calificarea examinatorului

Testarea psihologică efectuată de nespecialişti este extrem de dăunătoare, atât pentru subiect, cât şi pentru cel care a solicitat examinarea (organizatorul unui program de instruire care trebuie să trieze candidaţii, compania care doreşte să angajeze personal etc.), întrucât există riscuri mari să se ajungă la concluzii false. Din acest motiv, *La Comission Internationale de Reglementation de la Production et de la Vente des Tests*, la Congresul Asociaţiei Internaţionale de Psihologie Aplicată, în 1971, a decis să reglementeze distribuirea testelor psihologice astfel:

- **autorilor** li se cere să limiteze distribuirea testelor la persoanele care, prin formaţia lor şi prin respectul manifestat faţă de metodologia şi deontologia profesiei lor, sunt capabile să le folosească;
- **editorilor** li se cere să controleze distribuirea testelor, vânzându-le direct numai utilizatorilor calificaţi profesional;
- **utilizatorilor** li se cere să aplice testele numai conform indicaţiilor date de autorii acestora. Ei trebuie să respecte drepturile de autor, să nu reproducă, traducă, adapteze sau modifice un test fără permisiunea autorului şi a editorului. Persoanele a căror formaţie profesională este incompletă nu pot cumpăra şi utiliza teste decât prin intermediul şi sub supravegherea unei persoane care are calificarea necesară.

Aceste hotărâri sunt respectate de toate marile firme care produc, distribuie sau utilizează teste psihologice. De exemplu, *Etablissement d'Applications Psychotechniques* (EAP), o instituţie franceză care echipează cu teste psihologice cele mai mari centre de examinare psihologică din Franţa (SNCF, Citroen etc.) şi din lume (România, Polonia, Belgia, Olanda etc.) obligă persoanele care doresc să cumpere teste psihologice să prezinte o fotocopie a diplomei de licenţă în psihologie.

Numeroase state au acceptat ca dreptul de profesare a psihologiei să fie acordat numai pe baza unor standarde de competenţă, cum sunt titlul de doctor

în psihologie, diploma de master în psihologie sau diploma de absolvire a unui curs de specializare într-un domeniu al psihologiei. Diploma de absolvire a facultăţii de psihologie este adesea insuficientă.

De exemplu, *Guidelines for Computer-Based Tests and Interpretations* (APA, 1986) apreciază că au dreptul de utilizare a testelor psihologice doar "profesioniştii calificaţi" care, în plus, îndeplinesc şi următoarele condiţii (după Eyde, Kowal, 1987):

- au cunoştinţe asupra măsurării psihologice;
- cunosc istoricul testelor sau al chestionarelor pe care le utilizează;
- au experienţă în utilizarea testelor, astfel încât pot aprecia când există diferenţe de sex, vârstă sau cultură între rezultatele la un test;
- au cunoştinţe în domeniul în care aplică testele.

De exemplu, în cazul chestionarelor de personalitate, utilizatorul trebuie să aibă cunoştinţe de psihopatologie şi de teoria personalităţii.

Cataloagele care prezintă teste psihologice specifică şi nivelul de competenţă necesar cumpărării şi utilizării fiecărui test. În multe ţări, studenţii pot achiziţiona teste psihologice, doar dacă le prezintă distribuitorilor de teste semnătura persoanei responsabile de activitatea lor.

B. Calitatea instrumentelor şi a procedurilor de măsurare

Înainte de a utiliza un test, psihologul are obligaţia de a citi manualul testului, pentru a afla dacă:

- testul este adecvat scopului urmărit (de exemplu, dacă măsoară trăsătura care prezintă interes);
- testul este potrivit caracteristicilor subiectului /subiecţilor (sex, vârstă, nivel de studii etc.);
- calităţile testului (validitate, fidelitate, dificultate etc.) sunt acceptabile.

Este important de reţinut că felul în care acţionează un test depinde nu numai de calităţile sale, constatate în diverse experimente, ci şi de scopul pentru care este folosit şi de celelalte procedee de examinare psihologică utilizate împreună cu el. De exemplu, s-a constatat că, în cazul selecţiei profesionale, subiecţii au mai multă încredere în informaţiile furnizate de testele de aptitudini decât în cele obţinute din chestionarele de personalitate, ceea ce influenţează negativ sinceritatea răspunsurilor la itemii chestionarelor. S-a observat, de asemenea, că asocierea chestionarelor de personalitate cu teste de aptitudini şi cu un interviu creează o atitudine mai favorabilă faţă de examenul psihologic decât atunci când chestionarele de personalitate sunt însoţite doar de interviu (Fletcher, 1997).

Testarea psihologică necesită pregătiri pentru asigurarea calităţii bune a materialelor utilizate şi a condiţiilor de examinare.

Un test redactat neîngrijit, cu greşeli de ortografie, pătat sau cu figuri desenate neglijent le va face o impresie proastă subiecţilor, scăzându-le interesul pentru completarea sa ori, uneori, derutându-i. Performanţele la test vor fi, astfel, afectate.

Examenele psihologice improvizate în săli lipsite de mese, gălăgioase sau cu o luminozitate improprie vor conduce, de asemenea, la distorsionarea rezultatelor.

Este indicat ca examenul psihologic să fie anunţat din timp şi să se specifice cum trebuie să se pregătească subiecţii: să nu se prezinte la examen obosiţi, să nu consume alcool înaintea examinării, dacă folosesc ochelari, să îi aibă asupra lor etc.

C. Obţinerea consimţământului subiectului pentru examinarea psihologică

Psihologul are obligaţia să îi explice subiectului motivul examenului psihologic şi să obţină acordul acestuia de a se supune la test. Între persoana testată şi examinator se va încheia o convenţie scrisă, semnată de ambele persoane, prin care subiectul declară că a primit explicaţii ample despre scopul examinării, despre testele care vor fi utilizate şi despre modul în care vor fi folosite rezultatele, îşi dă consimţământul să fie examinat psihologic şi indică persoanele cărora le pot fi comunicate rezultatele examinării.

Dar, obţinerea consimţământului unei persoane pentru a participa la o examinare psihologică poate denatura rezultatele testelor, întrucât, cunoscând ceea ce se intenţionează să se măsoare, subiectul poate răspunde nesincer, din dorinţa de a crea o anumită imagine despre sine. Psihologul este pus uneori în situaţia de a opta între a obţine date corecte despre indivizii testaţi şi a respecta dreptul celor examinaţi de a fi informaţi despre test. În general, atunci când intruziunea testului în viaţa particulară este minoră şi se asigură confidenţialitatea răspunsurilor nu li se mai dau subiecţilor informaţii detaliate despre test.

D. Confidenţialitatea rezultatelor

Cu excepţia cazului când persoana examinată reprezintă un pericol pentru sine sau pentru alţii, informaţia furnizată de testele psihologice este confidenţială şi nu va fi comunicată altora fără consimţământul subiectului. Chiar şi atunci când subiectul este de acord ca rezultatele sale să fie cunoscute şi de alţii, rezultatele vor fi arătate doar medicului, avocatului sau psihologului care se ocupă de persoana respectivă.

Acceptând ideea că scorurile realizate la un test pot fi afectate de erori, uneori mari, care pot cauza interpretări neconforme cu realitatea, cu ocazia

comunicării rezultatelor trebuie să i se dea posibilitate subiectului să corecteze ceea ce el consideră că este o impresie nefavorabilă despre sine (Fletcher, 1997). Din acest motiv, în unele ţări, cum este Olanda, codul etic pentru psihologi prevede ca rezultatele testării să fie comunicate mai întâi subiectului.

În mod obişnuit trebuie evitată afişarea rezultatelor examenului psihologic. De exemplu, în cazul selecţiei profesionale se recomandă să se afişeze doar numele persoanelor declarate admise, fără alte date care ar putea constitui o violare a intimităţii subiecţilor. Cei nereuşiţi nici nu trebuie amintiţi. Acest lucru se impune mai ales atunci când sunt utilizate teste de inteligenţă sau de personalitate, întrucât denumirile acestora pot sugera interpretări care să îi lezeze pe subiecţii cu scoruri foarte mici (sau foarte mari, în funcţie de test).

E. Comunicarea rezultatelor testării

Niciodată rezultatele unui examen psihologic nu sunt comunicate sub forma lor brută, ci doar prelucrate şi interpretate. Obişnuit, ele sunt prezentate în cadrul unui *raport psihologic*, întocmit în conformitate cu solicitarea beneficiarului. Acest raport va fi redactat clar, fără ambiguităţi. Se va evita utilizarea termenilor tehnici, dificili de înţeles de nespecialişti. De asemenea, nu se va face presupunerea că beneficiarul cunoaşte testul şi îi sunt suficiente doar informaţii sumare. De exemplu, unui nespecialist trebuie să i se explice ce înseamnă faptul că o persoană a obţinut un coeficient de inteligenţă de 87 şi ce implicaţii ar putea avea aceasta. Datele comunicate trebuie să reflecte obiectiv situaţia, evitându-se observaţiile subiective şi părerile personale.

O deficienţă a multor rapoarte psihologice constă în faptul că ele descriu comportamente (de exemplu, "în timpul discuţiilor, subiectul gesticulează continuu"), fără a oferi comentarii asupra lor. Aceasta se întâmplă, de obicei, atunci când nici psihologul nu a putut interpreta faptele sau a considerat că explicaţiile nu prezintă interes pentru cel care citeşte raportul. Dar, simpla enumerare a observaţiilor făcute, fără comentarii, poate genera o mare varietate de interpretări şi poate conduce la formarea unei imagini greşite asupra subiectului. Este preferabil ca în raport să fie incluse toate explicaţiile pe care psihologul le consideră posibile, atrăgându-se atenţia că ele nu sunt certe.

Uneori, rezultatele obţinute la teste de o persoană sunt contradictorii. De exemplu, subiectul poate să apară la un test ca introvertit, iar la altul, ca sociabil. În asemenea situaţii este important ca psihologul să prezinte toate constatările făcute şi să încerce să găsească explicaţii.

F. Efectul discriminatoriu al testului asupra grupurilor minoritare

Aici termenul de "grup minoritar" este legat de contextul în care se aplică testul şi defineşte un grup, format în funcţie de o anumită variabilă (sex,

statut social, nivel socio-cultural, apartenenţă etnică etc.), care conţine mult mai puţine persoane decât celelalte grupe. De exemplu, la selecţiile profesionale pentru postul de secretară, faptul că numărul bărbaţilor care se prezintă este, de obicei, mult mai scăzut decât cel al femeilor, face ca bărbaţii să constituie un grup minoritar.

Problema discriminării produse de utilizarea unui test se pune mai ales atunci când testul respectiv este folosit în cadrul selecţiei profesionale şi se utilizează un singur prag de admisibilitate (cutoff score), adică o singură valoare cu care se compară scorurile subiecţilor pentru a se decide dacă sunt admişi sau respinşi. Dacă există grupuri minoritare care obţin scoruri mai mici decât alte grupuri, persoanele care fac parte din ele vor avea şanse mai mici de a fi selectate. Ele se vor considera victime ale discriminării. Realitatea nu este totdeauna aceasta. Să analizăm două situaţii în care se prezintă la selecţie două grupuri, unul cu scoruri mari la test şi celălalt cu scoruri mici.

(1) Să presupunem mai întâi că testul este construit astfel încât rezolvarea sa nu este influenţată de elemente care nu au nici o legătură cu variabila prezisă de test, numită criteriu. De exemplu, testul este utilizat la selecţia profesională pentru un post de casier; el se referă la deprinderea de a efectua rapid şi corect calcule aritmetice şi rezolvarea sa nu este dependentă de gradul de stăpânire a limbii. Criteriul utilizat fiind performanţa în muncă, dacă linia de regresie a criteriului faţă de test este aceeaşi pentru ambele grupuri, atunci din faptul că un grup obţine scoruri mai mici la test se deduce că el va avea rezultate mai slabe şi în muncă, deci este normal ca persoanele din grupul respectiv să fie respinse la selecţie. În acest caz testul nu are efect discriminatoriu.

(2) Se poate întâmpla, însă, ca cele două grupuri să fie la fel de potrivite pentru postul pentru care se face selecţia (performanţele lor în muncă sunt similare), dar liniile de regresie ale criteriului faţă de scorurile testului să difere de la un grup la altul. Astfel de situaţii apar atunci când persoanele dintr-un grup obţin rezultate slabe la test dintr-un motiv care în activitatea viitoare nu are importanţă sau poate fi compensat (de exemplu din cauza unui handicap de limbă ori cultură). Dacă se foloseşte acelaşi prag de admisibilitate pentru ambele grupuri, vor fi respinse unele persoane cu scoruri mici la test care ar fi putut obţine performanţe în muncă mai bune decât unele persoane admise. În asemenea cazuri se poate vorbi despre efectul discriminatoriu al testului, întrucât sunt defavorizaţi la selecţie cei din grupul cu scoruri mici la test. Pentru acest grup ar fi trebuit fixat un prag de admisibilitate mai scăzut.

CAPITOLUL 2
CÂTEVA NOŢIUNI CARE INTERVIN ÎN TEORIA TESTELOR PSIHOLOGICE

2.1. CONSTRUCTUL PSIHOLOGIC

Un **construct** este, în general, o idee "construită" de experţi dintr-un domeniu particular al ştiinţei, pentru a rezuma un grup de fenomene sau de obiecte şi pentru a fi utilizată într-un cadru ştiinţific (teoretic, metodologic sau aplicativ).

După Murphy şi Davidshofer (1991), orice construct are două proprietăţi:

- este o abstractizare a unor regularităţi din natură;
- nu este observabil direct, dar poate fi conectat cu entităţi sau evenimente concrete, observabile.

Constructele psihologice sunt evidenţiate, direct sau indirect, de comportament sau în urma conducerii unor experimente dedicate măsurării lor. Un construct psihologic, arată J.F. Binning şi G.V. Barett (Silva, 1993), este "o etichetă aplicată unei grupări de comportamente care covariază". Pentru *Standards for Educational and Psychological Tests* din 1985, constructul este "o noţiune teoretică derivată din cercetare şi din alte experienţe, care a fost construită pentru a explica patternuri de comportament observabile".

Constructele create pentru interpretarea rezultatelor testelor psihologice sunt identificate, de obicei, cu câte o clasă de răspunsuri (L. J. Cronbach, după Silva, 1993).

Unii autori consideră constructele ca fiind multidimensionale, formate din "componente" independente unele de altele, numite uneori "faţete", şi recomandă utilizarea analizei factoriale ca metodă pentru studierea acestora. De exemplu, în modelul elaborat de Guilford, constructul "creativitate" are drept componente: fluiditatea, flexibilitatea, originalitatea, elaborarea, senzitivitatea faţă de probleme şi redefinirea (Landau, 1979). Prezenţa componentelor ridică probleme în studiul relaţiei dintre construct, privit ca variabilă independentă, şi o variabilă dependentă prezisă de construct. O posibilitate de rezolvare este cotarea separată a componentelor, cumularea scorurilor şi cercetarea relaţiei dintre scorul total obţinut şi valorile variabilei dependente. Dar, uneori, componentele ar putea să prezică mai bine variabila dependentă decât constructul general. În asemenea situaţii se studiază separat relaţia dintre fiecare componentă şi variabila dependentă (Ozer, Reise, 1994).

Fiecare construct psihologic este fundamentat pe o teorie, care permite descrierea şi predicţia comportamentelor legate de el, în situaţii specifice.

Pentru elaborarea unui test care să îl măsoare, constructului i se asociază o variabilă cantitativă despre care se presupune că se află în relaţie crescătoare cu scorurile testului. Se formulează apoi o serie de afirmaţii referitoare la modul în care se comportă persoanele care au valori mari ale acestei variabile, comparativ cu cele care au valori scăzute. De exemplu, vrând să dezvolte un test de evaluare a motivaţiei indivizilor pentru propria realizare, McClelland şi colaboratorii săi au formulat următoarele descrieri de comportamente (Thorndike, Haagen, 1961):

- Persoanele care au o motivaţie puternică pentru realizare au rezultate şcolare bune.
- Motivaţia pentru realizare este mai mare dacă subiecţilor li se administrează în prealabil teste despre care li se spune că măsoară inteligenţa.
- Cei care au o motivaţie puternică pentru realizare completează mai mulţi itemi într-un test de viteză a cărui funcţie le-a fost explicată.
- Motivaţia pentru realizare este mai mare la copiii a căror familie le-a încurajat de timpuriu independenţa.

Fiecare dintre aceste descrieri este bazată pe un fel de "teorie a comportamentului uman". De exemplu, realizarea şcolară este văzută ca o combinaţie între aptitudinea şcolară şi efort. Dintre persoanele cu acelaşi nivel aptitudinal, cele care au o motivaţie mai mare pentru realizare depun un efort mai mare şi obţin rezultate şcolare mai bune (Thorndike, Haagen, 1961).

Comportamentele prin care autorul testului consideră că se manifestă constructul pot fi influenţate de diverse variabile, precum vârsta, sexul, educaţia primită etc. Pentru unele dintre acestea legătura cu constructul poate corespunde unei situaţii reale, observată în populaţia căreia i se adresează testul şi acceptată de teoria care stă la baza constructului. Pentru altele, însă, ea denotă că aceste comportamente nu sunt caracteristice constructului. De asemenea, comportamentele prin care este descris constructul măsurat de test se pot asocia cu comportamente prin care se manifestă alte constructe psihologice, evaluate prin teste existente.

În consecinţă, elaborarea unui test care să măsoare un anumit construct impune descrierea constructului în termeni comportamentali concreţi. Operaţia se numeşte *explicarea constructului* şi constă din următorii trei paşi (Murphy, Davidshofer, 1991):

(1) Se identifică acele comportamente care au legătură cu constructul măsurat de test.

(2) Se identifică alte constructe şi pentru fiecare se decide dacă are sau nu legătură cu constructul măsurat de test.

(3) Pentru fiecare construct se alcătuieşte o listă de comportamente prin care acesta se manifestă. Pentru fiecare comportament, pe baza relaţiilor dintre constructe, se decide dacă are sau nu legătură cu constructul măsurat de test.

Rezultă astfel un sistem de relaţii care înglobează un set de constructe şi comportamente, denumit de L. J. Cronbach, *reţea nomologică* ("nomologică" în sensul că ea are la bază declaraţii exprimate sub formă de legi). Unii autori au folosit termenul de "reţea stohastică", susţinând că relaţiile dintre componentele reţelei sunt mai degrabă probabilistice decât de tip lege.

O descriere amănunţită a constructului trebuie să conţină şi relaţiile acestuia cu cât mai multe variabile fiziologice, psihologice etc. De exemplu, dacă se urmăreşte realizarea unui test pentru evaluarea inteligenţei, se pot avea în vedere următoarele relaţii ale constructului "inteligenţă" cu alte variabile (Gronlund, 1967; Kline, 1993):

- Scorurile testului cresc cu vârsta (se presupune că inteligenţa creşte cu vârsta aproximativ până la 16 ani).
- Scorurile testului prezic succesul şcolar.
- Scorurile testului concordă cu aprecierile făcute de profesori despre inteligenţa elevilor.
- Scorurile testului corelează pozitiv cu scorurile unor teste aşa-numite "de inteligenţă".
- Dacă se efectuează analiza factorială asupra acestui test şi a altor teste care măsoară aptitudini, se va constata că testul de inteligenţă are o saturaţie mare în primul factor general.
- Scorurile testului discriminează între un grup de copii handicapaţi mintal şi unul format din copii capabili de performanţe superioare.
- Scorurile testului sunt influenţate doar puţin de procesul instrucţional urmat de subiect.
- Scorurile testului discriminează între diverse grupe ocupaţionale. Unele profesii, precum cea de matematician sau de funcţionar superior, necesită o inteligenţă suuperioară, în timp ce altele pot fi exercitate şi de persoane cu o inteligenţă redusă.

Fiecare dintre afirmaţiile formulate în legătură cu constructul pe care testul trebuie să îl măsoare urmează să fie verificată experimental. Dacă unele dintre ele nu se confirmă, trebuie revăzută teoria care stă la baza constructului, pentru a sesiza eventualele omisiuni sau neconcordanţe.

2.2. DOMENIUL DE CONŢINUT AL UNUI TEST

Prin **domeniu de conţinut al unui test** se înţelege "mulţimea tuturor comportamentelor care pot fi utilizate pentru a măsura atributul specific sau caracteristica la care se referă testul" (Murphy, Davidshofer, 1991).

Definirea domeniului de conţinut este asemănătoare celei de definire a unui construct. Autorul testului, pe baza ideii pe care şi-a format-o despre ceea ce vrea să măsoare, reţine acele comportamente care presupune că ar fi manifestări ale constructului, respectiv, în cazul domeniului de conţinut, expresii ale performanţei sau ale cunoştinţelor subiecţilor.

Deosebirea dintre un construct şi un domeniu de conţinut constă numai în gradul de abstractizare implicat şi adesea este dificil de făcut distincţie între ele (Guion, 1986).

Descrierea domeniului de conţinut trebuie în primul rând să precizeze limitele acestuia. Apoi, pentru fiecare componentă se apreciază importanţa. În cazul testelor destinate să măsoare rezultatele unui program de instruire (cunoştinţe sau deprinderi), descrierea domeniului de conţinut al testului se bazează pe analiza conţinutului şi a obiectivelor instrucţionale ale cursului urmat de persoanele care vor fi testate. Se încearcă să se includă în domeniul de conţinut aspectele pe care mai mulţi experţi le consideră importante.

Exemplu.

Un test de cunoştinţe de psihometrie, administrat studenţilor din anul III de la facultatea de psihologie, după predarea cursului de Psihodiagnostic, poate avea ca domeniu de conţinut "cunoaşterea noţiunilor de bază ale teoriei testelor, conform materiei predate la curs".

Descrierea domeniului de conţinut cuprinde:

a. limitele domeniului: se au în vedere cunoştinţele cuprinse în bibliografia indicată la curs (se anexează lista bibliografică);

b. componentele domeniului:

- probleme: analiza de itemi, validitatea testelor, fidelitatea testelor;
- aspecte tratate: definiţiile noţiunilor, metodele de cercetare, condiţiile de utilizare a metodelor de cercetare şi a formulelor de calcul;

c. importanţa relativă a componentelor domeniului: este indicată în tabelul 2.2.1. Procentele se referă la timpul alocat fiecărei probleme în cadrul cursurilor predate studenţilor.

Tabelul 2.2.1
Importanţa componentelor domeniului de conţinut pentru un test de cunoştinţe de psihometrie

Problema	Aspectul tratat		
	Definiţii	Metode	Condiţii de utilizare
Analiza de itemi	12%	14%	8%
Validitatea testelor	11%	17%	5%
Fidelitatea testelor	6%	19%	8%

Există, însă, multe domenii de conţinut care nu pot fi descrise detaliat după modelul prezentat. De exemplu, domeniul de conţinut al unui test de aptitudini cognitive poate fi format din itemi extraşi din teste cunoscute, care se consideră că evaluează aptitudinea respectivă. Într-o asemenea situaţie este dificil de precizat care sunt limitele domeniului de conţinut.

2.3. GRUPELE CONTRASTANTE

2.3.1. Definiţia grupelor contrastante

În multe cercetări efectuate în psihologie şi pedagogie se utilizează aşa-numitele *grupe contrastante* sau *grupe extreme*. Este vorba despre două loturi de subiecţi care, în privinţa uneia sau a mai multor variabile fixate de cercetător, au valori diferite: foarte mici într-o grupă, foarte mari în cealaltă.

Atunci când variabilele sunt măsurate pe scală de interval sau de raport, se poate da o definiţie matematică a grupelor contrastante. În acest scop, se realizează mai întâi o ordonare a variabilelor, după un anumit criteriu (de exemplu, se aşază în ordinea în care au fost înregistrate valorile lor). Apoi, pentru fiecare persoană, se transformă valorile variabilelor în cote standard z, utilizând mediile şi abaterile standard ale acestor variabile în mulţimea tuturor subiecţilor (mulţimea din care se formează cele două grupe extreme). Fiecărei persoane i se pune în corespondenţă un vector, ale cărui componente sunt egale cu cotele standard ale variabilelor, în ordinea stabilită.

DEFINIŢIA 2.3.1.1. Prin **centru de greutate** al unei mulţimi de vectori

$A=\{a_j=(a_{j1}, a_{j2},...,a_{jn}) | j=1,2,...,N\}$

se înţelege vectorul

$m=(m_1, m_2,...,m_n)$

are cărui componente m_i sunt egale cu mediile aritmetice ale valorilor componentelor corespunzătoare, în mulţimea dată:

$m_i=(a_{1i}+a_{2i}+...+a_{Ni}) / N$

$i=1,2,...,n$.

DEFINIŢIA 2.3.1.2. (Albu, Pitariu, 1995c) Două grupe de subiecţi se numesc **contrastante, la un prag de probabilitate fixat,** p, dacă sunt îndeplinite următoarele două condiţii:

a. Centrul de greutate al vectorilor corespunzători persoanelor unei grupe are toate componentele negative, semnificativ diferite de zero la pragul p, iar centrul de greutate al vectorilor corespunzători subiecţilor din cealaltă grupă are toate componentele pozitive, semnificativ diferite de zero la pragul p.

b. Componentele corespunzătoare ale celor două centre de greutate diferă semnificativ între ele, la pragul p fixat.

Observaţii

1. Prima condiţie din definiţia 2.3.1.2 asigură faptul că în grupa "bună" mediile valorilor variabilelor sunt superioare mediilor întregului lot, iar în grupa "slabă", ele sunt inferioare mediilor întregului eşantion.

2. Cea de a doua condiţie impune ca pentru fiecare variabilă mediile celor două grupe să fie semnificativ diferite între ele.

Fie (C_1, C_2) şi (C_3, C_4) două perechi de grupe contrastante în sensul definiţiei 2.3.1.2, grupa C_i având centrul de greutate

$$c_i = (x_{i1}, x_{i2},...,x_{in}), \quad i=1,2,3,4.$$

C_1 şi C_3 sunt grupele "slabe", iar C_2 şi C_4 sunt grupele "bune".

DEFINIŢIA 2.3.1.3. Se spune că elementele perechii (C_1, C_2) **contrastează mai puternic** decât elementele perechii (C_3, C_4) dacă are loc relaţia:

$$\min \{|x_{1j}-x_{2j}|: j=1,2,...,n\} > \min \{|x_{3j}-x_{4j}|: j=1,2,...,n\} \qquad (2.3.1.1)$$

Se spune că elementele celor două perechi **contrastează la fel de puternic** dacă are loc relaţia:

$$\min \{|x_{1j}-x_{2j}|: j=1,2,...,n\} = \min \{|x_{3j}-x_{4j}|: j=1,2,...,n\} \qquad (2.3.1.2)$$

2.3.2. Formarea grupelor contrastante

Modul de formare a grupelor contrastante este dependent de numărul variabilelor alese şi de scala pe care acestea sunt măsurate.

În cazul unei singure variabile dihotomice, cum este sexul, grupele contrastante corespund celor două valori ale variabilei.

Dacă pentru formarea grupelor contrastante se utilizează o singură variabilă, măsurată pe scală ordinală, de interval sau de raport, atunci se începe prin a ierarhiza persoanele pe baza valorilor acestei variabile. În prima grupă, denumită de obicei "grupa slabă", vor fi incluse persoanele cu valorile cele mai mici ale variabilei, iar în cealaltă grupă, denumită "grupa bună", persoanele cu valorile cele mai mari. În literatura de specialitate există recomandarea ca cele două grupe să conţină câte 27% din totalul subiecţilor. Dar, din motive practice (se poate întâmpla ca mai multe persoane să aibă aceeaşi valoare a variabilei) se acceptă ca procentul celor incluşi în fiecare grupă să fie cuprins între 25% şi 33%.

Uneori pentru formarea grupelor contrastante este necesar să se ia în considerare mai multe variabile, fiecare fiind măsurată pe scală ordinală, de interval sau de raport. Aşa se întâmplă, de pildă, când grupele trebuie să conţină persoane "cu succes profesional" şi, respectiv, persoane "cu insucces profesional", iar performanţa profesională se exprimă prin valorile mai multor variabile. În asemenea situaţii, se procedează la ierarhizarea persoanelor după valorile fiecărei variabile şi la agregarea clasamentelor astfel formate. Din ierarhia rezultată se aleg cei aflaţi la extreme, ca şi în cazul descris mai sus.

În continuare va fi descris un algoritm de formare a grupelor contrastante aplicabil în cazul când numărul variabilelor utilizate este cel puţin egal cu 2 şi când variabilele sunt măsurate pe scală de interval.

Algoritmul construieşte, la fiecare pas, o partiţie a mulţimii de subiecţi aplicând analiza de clusteri (Everitt, Dunn, 1983); în acest fel, clasele formate sunt omogene. Dintre clasele partiţiei se determină cele care contrastează cel mai puternic.

Fie $S=\{s_1, s_2,...,s_N\}$ mulţimea vectorilor n-dimensionali corespunzători celor N subiecţi pentru care se cunosc cotele standard la n variabile măsurate pe scală de interval sau de raport. Algoritmul care determină grupele contrastante ale mulţimii S acţionează astfel:

Pasul 1:

Se stabileşte volumul minim (numărul minim de persoane) al grupelor contrastante, N_{min}, şi pragul de probabilitate utilizat la compararea mediilor, p.

Se construieşte pe mulţimea S partiţia discretă

$$P=\{\{s_1\}, \{s_2\},...\{s_N\}\}.$$

Centrul de greutate al fiecărei clase $\{s_i\}$ este vectorul s_i, $i=1,2,...,N$.

Pasul 2:

Se formează mulţimile:

P_- = mulţimea claselor partiţiei P care au toate componentele centrului de greutate negative;

P_+ = mulţimea claselor partiţiei P care au toate componentele centrului de greutate pozitive.

Dacă $P_-=\emptyset$ sau $P_+=\emptyset$ atunci se trece la pasul 4.

Pasul 3:

Se extrag din mulţimile P_- şi P_+ clasele care conţin cel puţin N_{min} elemente şi se formează mulţimile $P_{-,min}$ şi $P_{+,min}$:

$P_{-,min}=\{C_- : C_-\in P_-, |C_-|\geq N_{min}\}$

$P_{+,min}=\{C_+ : C_+\in P_+, |C_+|\geq N_{min}\}$

(S-a notat cu $|C_-|$ şi $|C_+|$ numărul elementelor mulţimii C_-, respectiv al mulţimii C_+.)

Dacă $P_{-,min}=\emptyset$ sau $P_{+,min}=\emptyset$ atunci se trece la pasul 4, în caz contrar se trece la pasul 5.

Pasul 4:

Se reunesc două clase din partiţia P efectuând un pas al algoritmului analizei de clusteri. Se formează astfel o nouă partiţie P pe mulţimea S.

Dacă partiţia P are o singură clasă se trece la pasul 7.

În caz contrar, pentru clasa formată la acest pas se determină centrul de greutate şi se trece la pasul 2.

Pasul 5:

Se determină mulţimea A alcătuită din toate perechile posibile formate dintr-o clasă din $P_{-,min}$ şi o clasă din $P_{+,min}$ care îndeplinesc condiţiile definiţiei 2.3.1.2:

$$A=\{(C_-,C_+) : C_-\in P_{-,min}, C_+\in P_{+,min}, C_- \text{ şi } C_+ \text{ contrastante la pragul } p\}$$

Dacă $A=\emptyset$ atunci se trece la pasul 4, în caz contrar se trece la pasul 6.

Pasul 6:

Se defineşte în mulţimea A relaţia de preordine total $\geq$ astfel:

$(C_-,C_+), (G_-,G_+) \in A$:

$(C_-,C_+) \geq (G_-,G_+)$ <=> elementele perechii (C_-,C_+) contrastează mai puternic decât elementele perechii (G_-,G_+) sau elementele perechii (C_-,C_+) contrastează la fel de puternic ca şi elementele perechii (G_-,G_+).

Se determină perechile $(C_-,C_+) \in A$ pentru care

$(C_-,C_+) \geq (G_-,G_+)$ oricare ar fi o altă pereche $(G_-,G_+) \in A$.

Fiecare din aceste perechi este o soluţie a algoritmului.

Se trece la pasul 4.

Pasul 7: STOP

Atunci când algoritmul furnizează mai multe soluţii, se va alege aceea în care volumul fiecărei grupe este apropiat de valoarea 0.27x*N*.

Dacă algoritmul nu găseşte soluţii, se pot forma grupe contrastante prin metoda descrisă anterior, bazată pe agregarea clasamentelor corespunzătoare celor *n* variabile.

Exemplu

S-a considerat o mulţime de 27 de persoane pentru care se cunosc cotele realizate la două scale necorelate liniar între ele (r=0.095). Fiecărei persoane i s-a ataşat un vector bidimensional, reprezentând scorurile standard la cele două scale.

Pentru început s-au agregat ierarhiile produse de scorurile scalelor însumând, pentru fiecare persoană, scorurile standard şi ordonând subiecţii după scorul total. S-au inclus în fiecare din grupele contrastante aproximativ un sfert din totalul persoanelor (mai precis, 25.926%). Punctele corespunzătoare subiecţilor din cele două clase sunt prezentate în figura 2.3.2.1.

Se constată că în clasa "bună" sunt cuprinse şi persoane care la o scală au scoruri inferioare mediei lotului. De asemenea, în grupa "slabă" sunt incluse şi persoane care la o scală au realizat scoruri superioare mediei lotului.

S-a aplicat pentru acest lot şi algoritmul prezentat mai sus. La pasul 20 grupele contrastante au frecvenţe apropiate celor din exemplul anterior. Dar, de această dată, toţi subiecţii din grupa "bună" au scoruri superioare mediei lotului, la ambele scale; de asemenea, toate persoanele din grupa "slabă" au scoruri inferioare mediei lotului, la ambele scale (figura 2.3.2.2).

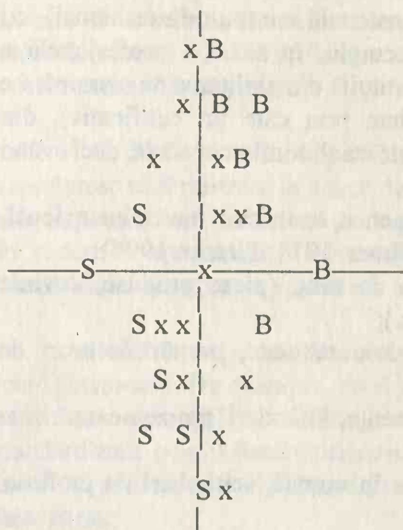


Figura 2.3.2.1

Formarea grupelor contrastante prin ierarhizarea persoanelor în funcţie de scorul total

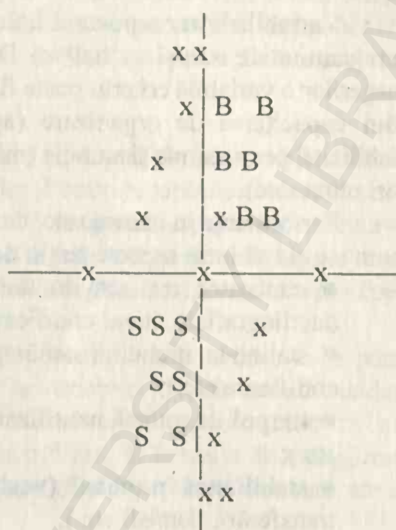


Figura 2.3.2.2

Formarea grupelor contrastante prin aplicarea algoritmului descris mai sus

Legendă:
B = persoană din grupa "bună"
S = persoană din grupa "slabă"
x = persoană care nu face parte din grupele contrastante

2.4. CRITERIUL

În teoria testelor psihologice, prin **criteriu** se înţelege "o variabilă pe care încercăm să o prevedem în general cu ajutorul testelor" (Grand dictionnaire de la psychologie, 1994) sau "o măsură acceptată a comportamentului evaluat de test" (Corsini, 1994). În aprecierea personalului, criteriul este o "măsură a nivelului de performanţă exprimată în termeni cantitativi, bazată pe o descriere completă a muncii prestate" (Corsini, 1994).

Criteriul este ecologic prin natura sa. El se referă la relaţia dintre persoane şi mediul lor (Barclay, 1968).

Variabila criteriu poate fi unidimensională sau multidimensională, cu valori cantitative sau/şi calitative. De exemplu, în selecţia profesională a managerilor o variabilă criteriu poate fi constituită din abilitatea de comunicare şi din capacitatea de organizare (apreciate prin câte un calificativ), din sociabilitate, persuasiune şi ambiţie (măsurate cu ajutorul unor scale, deci având valori numerice).

Performanţa în muncă este, de asemenea, un criteriu multidimensional. În compoziţia sa intră aspecte ce ţin de (Gilmer, 1971; Pitariu, 1994):

- cantitatea realizată în unitatea de timp (piese produse, cuvinte dactilografiate, itemi codificaţi etc.);
- calitatea muncii (număr de piese rebutate, număr de erori de codificare etc.);
- timpul de muncă neutilizat (absenţe, întârzieri, pauze neautorizate etc.);
- stabilitatea în muncă (vechimea în muncă, schimbări de profesie, transferări, demisii etc.);
- costul necesar pentru instruire (timpul de instruire pentru realizarea unei producţii standard, costul materialului stricat în timpul instruirii, viteza de avansare etc.);
- cunoştinţele profesionale;
- satisfacţia în muncă (număr de revendicări, de sugestii, examinări psihiatrice etc.);
- evaluările făcute de superiori sau de colegi.

Se recurge adesea la verificarea corectitudinii unui test, în cadrul operaţiei de validare a testului, prin compararea scorurilor sale cu măsurile unui criteriu, atunci când testul trebuie să înlocuiască variabila criteriu, fie din cauză că valorile variabilei criteriu nu sunt cunoscute în momentul testării, fie pentru că administrarea testului este mai puţin costisitoare decât măsurarea variabilei criteriu.

Pentru a face posibilă aprecierea corectă a calităţilor testului cercetat, criteriul trebuie să îndeplinească următoarele condiţii (adaptare după Greuter, Algera, 1989 şi Corsini, 1994):

A. Să fie *relevant* pentru activitatea sau caracteristica la care se referă, adică ordinea subiecţilor determinată pe baza valorilor criteriului să corespundă ordinii reale a acestora în privinţa succesului obţinut în activitatea respectivă sau a mărimii caracteristicii măsurate. Pentru aceasta este necesar ca variabila criteriu să ia valori într-o mulţime total ordonată ori să realizeze o măsurare pe scală ordinală, de interval sau de raport. De obicei se preferă criteriile cu valori cantitative.

B. Să fie *fidel*, adică să concorde cu diverse evaluări ale performanţelor la care el se referă, făcute la momente diferite şi/sau cu alte instrumente de măsură, similare lui. Verificarea acestei condiţii se poate face cu ajutorul coeficientului de corelaţie.

C. Să fie *practic*, adică să nu coste mult şi să fie acceptabil pentru cei care doresc să îl utilizeze la luarea deciziilor. Pentru îndeplinirea acestei condiţii este important ca în criteriu să fie înglobate principalele aspecte care sunt avute în vedere la aprecierea performanţei sau a mărimii caracteristicii măsurate. Identificarea acestora se poate face apelând la supervizori, respectiv la experţi.

D. Să fie exprimat în aceleaşi unităţi sau în unităţi comparabile pentru toate persoanele. De exemplu, dacă valorile variabilei criteriu sunt dependente de vârsta subiecţilor, atunci aceste valori vor fi transformate în cote standardizate, pentru fiecare persoană fiind utilizate în formula de transformare media şi abaterea standard corespunzătoare grupei de vârstă din care aceasta face parte.

E. Atunci când criteriul este multidimensional, valoarea criteriului pentru fiecare persoană să reflecte atât importanţa diverselor componente, cât şi nivelul individului la fiecare dintre ele.

Anastasi (1954) enumeră câteva criterii utilizate în practica psihologică pentru analizarea validităţii testelor care măsoară constructe. Acestea sunt:

a. *Vârsta*: se foloseşte drept criteriu pentru testele de inteligenţă, întrucât se consideră că scorurile acestora trebuie să crească pe parcursul copilăriei, până la maturitate. Nu este un criteriu bun pentru testele care măsoară funcţiuni psihice invariabile sau puţin variabile în timp, cum sunt variabilele de personalitate.

b. *Cunoştinţele şcolare*, exprimate prin note şcolare, rezultate la teste de cunoştinţe ori evaluări făcute de profesori asupra "inteligenţei" elevilor: sunt adesea criterii pentru testele care măsoară aptitudinea şcolară.

c. *Performanţele la diverse programe de instruire specială*, cum sunt cursurile de dactilografie, programare, muzică etc.: furnizează criterii pentru teste de aptitudini speciale.

d. *Performanţa profesională* serveşte uneori drept criteriu pentru teste de personalitate.

e. *Evaluările făcute de profesori, instructori sau superiori* sunt folosite mai mult pentru a obţine informaţii despre alte criterii, cum ar fi cunoştinţele şcolare, performanţa la diverse cursuri, reuşita profesională. Au rol de criteriu

pentru unele teste care măsoară trăsături psihice, precum anxietatea, onestitatea, originalitatea etc.

f. *Rezultatele altor teste* se utilizează ca şi criteriu atunci când se urmăreşte să se construiască un test care să măsoare acelaşi lucru ca şi criteriul dar să fie mai uşor de administrat sau/şi de cotat ori să fie mai ieftin. De exemplu, scala Stanford-Binet a servit drept criteriu pentru multe teste de inteligenţă.

g. *Grupele contrastante* reprezintă de obicei un criteriu multidimensional, care cuprinde aspecte diverse din viaţa zilnică. De exemplu, pentru un test de inteligenţă grupele contrastante pot fi constituite dintr-o clasă de copii proveniţi de la o şcoală de debili mintali şi o clasă de elevi de aceeaşi vârstă dintr-o şcoală obişnuită. Criteriul pe baza căruia s-au format grupele este alcătuit din ansamblul de factori care au determinat cuprinderea copiilor în şcoala ajutătoare. Grupele contrastante se folosesc frecvent ca şi criteriu pentru testele de personalitate. De exemplu, pentru evaluarea unor trăsături legate de activitatea socială se poate forma o grupă de studenţi cu multiple preocupări extraşcolare şi una cu studenţi fără asemenea preocupări.

Criteriile utilizate în cercetările asupra comportamentului pot fi clasificate din două puncte de vedere, ambele permiţând plasarea criteriilor de-a lungul unei axe (Barclay, 1968). Astfel:

- după modul de obţinere a valorilor lor, la un pol se plasează criteriile *subiective* (valorile sunt furnizate de subiect), iar la celălalt, cele *obiective* (valorile sunt obţinute de la alte persoane sau cu ajutorul unor instrumente de măsură);
- după gradul în care valorile lor sunt dependente de contextul în care se află persoana evaluată, la un pol se află criteriile *individuale* (valorile sunt independente de contextul în care se află persoana), iar la celălalt, cele *de grup* (comportamentele sunt puternic influenţate de grupul din care face parte persoana evaluată cu ajutorul criteriului).

De exemplu, criterii subiective individuale sunt afirmaţiile făcute de subiect despre nivelul său de aspiraţie şi despre valorile sale spirituale, în timp ce criterii subiective de grup sunt relatările subiectului despre relaţiile sale cu alte persoane, despre recompensele şi pedepsele pe care le-a primit. Criterii obiective individuale sunt tehnicile Q-sort, iar dintre criteriile obiective de grup fac parte evaluările realizate de profesori şi observaţiile sistematice asupra comportamentului.

Adesea, pentru un acelaşi test se găsesc mai multe criterii, fiecare referindu-se la un alt aspect. Aşa se întâmplă, de pildă, când testul trebuie să prevadă reuşita profesională.

În asemenea cazuri există două posibilităţi de a acţiona pentru a le lua în considerare pe toate:

a. se combină criteriile într-un *criteriu compus* sau *global*; pe baza scorurilor criteriilor se calculează pentru fiecare persoană o cotă unică;

b. se reunesc criteriile într-un *criteriu multiplu*: fiecare criteriu are o valoare particulară şi este utilizat separat în aprecierea testului.

Prima variantă de lucru este mai convenabilă în privinţa prelucrării şi a interpretării datelor, întrucât fiecare persoană are un singur scor, care permite ierarhizarea candidaţilor. Pentru aplicarea sa, este necesar, însă, ca scorurile criteriilor să fie exprimate în aceleaşi unităţi de măsură: în scoruri standardizate, în ranguri, în unităţi de timp sau în costuri.

Exemplu.

Pentru munca de editare a textelor la calculator se pot fixa două variabile criteriu:

- numărul de semne introduse într-o oră;
- numărul de erori comise la introducerea acestor semne.

Cu ajutorul acestor criterii se poate forma un criteriu global, care exprimă timpul necesar pentru a introduce corect un număr fixat de semne, *N*. Valoarea criteriului global, pentru fiecare persoană, se calculează însumând timpul în care sunt introduse cele *N* semne cu timpul necesar pentru depistarea erorilor comise şi corectarea lor. Prin combinarea criteriilor într-un criteriu global şi ierarhizarea persoanelor pe baza scorului astfel calculat se rezolvă o problemă de agregare a mai multor clasamente: ierarhiile de pornire (corespunzătoare scorurilor criteriilor) se înlocuiesc cu un singur clasament.

În practica selecţiei profesionale se obişnuieşte să se combine liniar, ponderat sau neponderat, scorurile criteriilor pentru a obţine cota globală. În privinţa modului de stabilire a ponderilor, părerile sunt diverse. De exemplu, H. Toops (Brogden, Taylor, 1949) consideră că ponderile trebuie să reflecte importanţa fiecărui criteriu, apreciată de specialişti, în timp ce M.L. Blum şi J.C. Naylor (1968) propun ca ponderile să fie proporţionale cu fidelitatea criteriilor.

Procedeele de agregare a clasamentelor care însumează scorurile ponderate ale criteriilor prezintă un neajuns atunci când ierarhiile corespunzătoare criteriilor nu concordă între ele. Astfel, este posibil ca pe primele locuri ale ierarhiei finale să ajungă persoane care au obţinut scoruri modeste la mai multe criterii, dacă ele s-au aflat pe locuri fruntaşe în clasamentele corespunzătoare criteriilor cărora li se atribuie ponderi mari.

Acest inconvenient poate avea urmări grave atunci când criteriile nu au un caracter compensatoriu. Dacă existenţa unor cunoştinţe sau deprinderi este absolut necesară desfăşurării cu succes a activităţii pentru care se face selecţia

şi nu poate fi compensată prin alte calităţi ale persoanei, de asemenea necesare postului, aceste cunoştinţe sau deprinderi trebuie să constituie un criteriu care nu va putea fi, sub nici o formă, inclus într-un criteriu global (Landy, 1985). În asemenea situaţii trebuie să se utilizeze un criteriu multiplu.

2.5. VALIDITATEA DE ASPECT A TESTELOR

Orice test psihologic sau educaţional care este destinat utilizării în practică trebuie să fie perceput de cei examinaţi cu el ca fiind adecvat domeniului în care este aplicat. Această calitate a testelor este denumită **validitate de aspect** (C. I. Mosier, după Nevo, 1993).

O definiţie operaţională a validităţii de aspect a fost formulată de L. Guttman şi S. Shye (Nevo, 1993) astfel: "Se face o măsurare a validităţii de aspect atunci când o persoană testată /un utilizator neprofesionist /un individ interesat, utilizând o tehnică absolută /relativă apreciază un item / un test / o baterie de teste drept foarte potrivit /.../ nepotrivit pentru utilizarea care îl interesează".

Trebuie observat că aprecierea validităţii de aspect nu este realizată de experţi şi pentru măsurarea sa nu se folosesc metode psihometrice. Dacă aprecierea validităţii de aspect este făcută de persoanele testate, ele trebuie să îşi spună părerea imediat după terminarea testului, când impresiile sunt încă proaspete. Validitatea de aspect nu este inclusă printre celelalte tipuri de validitate (*Standards for Educational and Psychological Tests*, 1974) şi nici nu le poate înlocui.

2.6. VARIABILE MODERATOARE, VARIABILE MEDIATOARE

Noţiunea de **variabilă moderatoare** are mai multe accepţiuni.

În termeni generali, o variabilă moderatoare este o variabilă calitativă (de exemplu, sexul, rasa, clasa socială) sau cantitativă (de exemplu, vârsta, anxietatea) care afectează direcţia şi/sau tăria relaţiei dintre o variabilă independentă şi o variabilă dependentă (Baron, Kenny, 1986).

Unii autori înţeleg prin variabilă moderatoare "o interacţiune între mai multe variabile care influenţează corelaţiile uneia dintre variabile cu alte date" (Mischel, 1968).

Procedeul statistic utilizat pentru a verifica dacă o variabilă independentă acţionează sau nu ca o variabilă moderatoare este determinat de felul în care au fost măsurate variabila independentă şi variabila moderatoare. Se face distincţie între cazul când variabilele sunt *categoriale*, adică discrete, măsurate pe scală nominală sau ordinală, şi cel în care ele sunt *continue*, măsurate pe scală de interval sau de raport.

Atunci când variabila dependentă este continuă sunt posibile următoarele situaţii (Hofstee, Smid, 1986; Baron, Kenny, 1986):

A. Variabila independentă şi variabila moderatoare sunt ambele categoriale.

În acest caz se aplică analiza de varianţă bidimensională pentru a cerceta efectul variabilei independente şi al variabilei despre care se presupune că este moderatoare, asupra variabilei dependente. Existenţa efectului moderator este indicată de prezenţa interdependenţei dintre variabila independentă şi cea moderatoare.

B. Una dintre cele două variabile (independentă şi moderatoare) este continuă, iar cealaltă categorială.

În acest caz se determină ecuaţia de regresie a variabilei dependente faţă de variabila continuă, pentru fiecare valoare a variabilei categoriale. Dacă există un efect moderator, atunci dreptele de regresie nu sunt paralele.

C. Variabila moderatoare şi variabila independentă sunt ambele continue.

Dacă se poate presupune că pentru variabila moderatoare există intervale de valori disjuncte consecutive, astfel încât relaţia dintre variabila independentă şi cea dependentă este aceeaşi în cadrul fiecărui interval dar se schimbă de la un interval la altul, atunci se poate proceda ca în cazul B, considerând fiecare interval ca o singură valoare.

Dacă, însă, se presupune că relaţia dintre variabila independentă şi cea dependentă variază continuu în funcţie de valorile variabilei moderatoare, atunci se determină ecuaţia de regresie a variabilei dependente Y faţă de variabila independentă X şi faţă de variabila moderatoare Z, sub forma:

$$Y = aX + bZ + cXZ + d \qquad (2.6.1)$$

Prezenţa efectului moderator este semnalată de valoarea nulă a coeficientul c.

Alţi autori (MacCallum, Mar, 1995) consideră că există un efect moderator dacă diferenţa dintre pătratele coeficienţilor de corelaţie multiplă corespunzători ecuaţiilor de regresie

$$Y = aX + bZ + cXZ + d \quad (2.6.2)$$
şi
$$Y = eX + fZ + g \quad (2.6.3)$$
este semnificativă statistic.

O variabilă *U* acţionează ca **mediator** între variabila independentă *X* şi cea dependentă *Y* dacă sunt îndeplinite următoarele condiţii (Baron, Kenny, 1986):

- variaţia valorilor variabilei independente provoacă variaţia valorilor variabilei mediatoare;
- variaţia valorilor variabilei mediatoare determină variaţia valorilor variabilei dependente;
- relaţia dintre variabila independentă şi cea dependentă este semnificativă, dar devine slabă sau chiar inexistentă atunci când sunt controlate relaţiile dintre variabila independentă şi cea mediatoare şi dintre variabila mediatoare şi cea dependentă.

Procedeul statistic aplicat pentru verificarea existenţei efectului mediator constă în construirea ecuaţiilor de regresie
$$U_* = dX_* \quad (2.6.4)$$
$$Y_* = cX_* \quad (2.6.5)$$
$$Y_* = aX_* + bU_* \quad (2.6.6)$$
unde X_*, Y_* şi U reprezintă variabilele standardizate corespunzătoare variabilelor *X*, *Y* şi *U* (adică, variabile având media 0 şi abaterea standard 1).

Dacă există efect mediator, atunci coeficienţii *d*, *c* şi *b* sunt semnificativi statistic, iar coeficientul *a* este mai mic decât *c*. Medierea perfectă apare atunci când, variabila mediatoare fiind controlată, variabila independentă nu are nici un efect asupra variabilei dependente.

În general, în cercetările psihologice mediatorii sunt diverse caracteristici ale persoanelor, care transformă variabila de intrare în variabila de ieşire.

Variabilele moderatoare sunt utilizate atunci când relaţia dintre variabila independentă şi cea dependentă este slabă, ea modificându-se de la o subpopulaţie la alta. De exemplu, gradul în care o trăsătură psihică este relevantă pentru subiecţi acţionează ca variabilă moderatoare dacă trăsătura trebuie să prezică un anumit comportament.

Variabilele mediatoare se folosesc atunci când există o relaţie strânsă între variabila independentă şi cea dependentă.

Uneori, în spatele unei variabile moderatoare se află o variabilă mediatoare.

Exemplu.

O cercetare privind anxietatea faţă de calculator a studenţilor psihologi (Pitariu, Albu, 1995) a evidenţiat faptul că între extraversiune şi convingerea studenţilor că nu pot învăţa să lucreze la calculator (unul dintre factorii anxietăţii faţă de calculator) intervine, ca variabilă mediatoare, modul în care subiecţii îşi apreciază cunoştinţele de informatică (nota pe care şi-o acordă). Variabila mediatoare corelează pozitiv cu extraversiunea şi negativ cu factorul anxietăţii faţă de calculator. Pe de altă parte, autoaprecierea subiecţilor corelează liniar pozitiv cu nivelul real al cunoştinţelor de informatică ale acestora, care este strâns legat de profilul liceului absolvit (liceu de matematică-fizică, de filologie, de arte etc.). Dacă cercetarea s-ar fi restrâns la mai puţine variabile, neluând în considerare nivelul cunoştinţelor de informatică -autoapreciat sau apreciat obiectiv- s-ar fi constatat că profilul liceului absolvit joacă rol de variabilă moderatoare între extraversiune şi anxietatea faţă de calculator.

În unele situaţii, o variabilă poate acţiona atât ca moderator, cât şi ca mediator. De exemplu, auto-eficienţa (care a fost definită de A. Bandura ca "aprecierea făcută de individ asupra capacităţii proprii de a organiza şi executa acţiunile necesare pentru a atinge o anumită performanţă") funcţionează ca:

- mediator între situaţia de lucru -apreciată din punctul de vedere al complexităţii lucrului şi al controlului deţinut de subiect- şi iniţiativa personală; aceasta înseamnă că atât controlul pe care îl are subiectul asupra muncii cât şi complexitatea lucrului au un efect direct asupra auto-eficienţei sale şi aceasta are efect direct asupra iniţiativei personale;
- moderator în relaţia dintre situaţia de lucru şi iniţiativa personală; relaţia dintre controlul deţinut de subiect şi complexitatea lucrului, pe de o parte, şi iniţiativa personală, pe de altă parte, este dependentă de nivelul de auto-eficienţă (Speier, Frese, 1997).

CAPITOLUL 3
VECHI ŞI NOU ÎN TESTAREA PSIHOLOGICĂ

3.1. SCURT ISTORIC AL TESTELOR PSIHOLOGICE

Originea cuvântului "test" este incertă. Ipoteza cea mai probabilă este că el provine din cuvântul latin "testa", care denumeşte un fel de creuzet folosit la separarea substanţelor nobile de celelalte. O altă posibilitate ar fi ca la originea sa să stea cuvântul latin "testatio", care înseamnă dovadă, depunere ca martor. Testul ar desemna, în acest caz, o ocazie de a face dovada capacităţilor proprii.

Începuturile testării psihologice se pierd în antichitate. Cu patru mii de ani în urmă, în Imperiul Chinez se utilizau teste de cunoştinţe pentru selecţia funcţionarilor publici.

În Evul Mediu, universităţile europene foloseau o serie de teste pentru promovarea studenţilor în ani superiori de studii.

Testele psihologice construite pe baze ştiinţifice datează doar de aproximativ o sută de ani. Apariţia lor este legată de constituirea, în a doua jumătate a secolului trecut, a psihologiei ştiinţifice şi, în cadrul ei, a metodelor psihometrice. O contribuţie importantă în acest domeniu şi-au adus-o, în Germania, Ernst Weber, Gustav Fechner şi Hermann Helmholz, prin preocuparea lor pentru măsurarea precisă a fenomenelor psihice, pentru punerea la punct a tehnicilor de analiză a reacţiilor şi pentru utilizarea metodelor matematice la interpretarea rezultatelor. În acelaşi timp, în Anglia, Sir Francis Galton (1822-1911), elev al lui Darwin, şi-a orientat cercetările spre măsurarea diferenţelor individuale. În cartea sa *Hereditary Genius*, editată în anul 1869, el formulează ideea că toate aptitudinile umane sunt distribuite conform curbei lui Gauss şi, deci, indivizii pot fi clasificaţi în funcţie de proprietăţile cunoscute ale legii repartiţiei normale.

În anul 1879, Wilhelm Wundt (1832-1920) a înfiinţat la Leipzig primul laborator de psihologie experimentală, iar Galton a publicat în revista "Brain" articolul *Psychometric Experiments*, în care afirmă necesitatea utilizării metodelor cantitative în ştiinţele umane şi introduce termenul **psihometrie**, pe care îl defineşte ca "arta de a impune operaţiilor spiritului măsura şi numărul, aşa cum se întâmplă, de exemplu, când se determină timpul de reacţie al persoanelor" (Meuris, 1985).

Elev al lui Wundt şi, apoi, al lui Galton, psihologul american James Mc Keen Cattell (1860-1944) publică în 1890, în revista engleză "Mind" articolul

intitulat *Mental tests and measurements*, în care apare, pentru prima dată, termenul **test**, în sensul de "serie de probe psihologice utilizate pentru a studia diferenţele individuale". Noţiunea de "test" este, deci, asociată cu cea de "măsură".

În 1897, Hermann Ebbinghaus elaborează un test de evaluare a capacitaţii mintale pentru elevi, legând astfel psihometria de preocupările pedagogice. Puţin mai târziu, în 1901, Ugo Pizzoli inaugurează folosirea testelor în orientarea profesională.

Dar, încă de la începutul aplicării metodei testelor au apărut şi primele critici la adresa sa. Observaţia făcută de C. Wissler, în 1901, la Universitatea Columbia, că notele şcolare corelează foarte puţin cu rezultatele testelor mintale a generat o atitudine de scepticism vizavi de utilitatea testelor.

Criticând faptul că majoritatea testelor existente evaluează funcţii psihice inferioare, Alfred Binet şi Victor Henri propun să se elaboreze teste pentru măsurarea unor "fenomene superioare ale spiritului, precum memoria, atenţia, imaginaţia, orientarea ideilor". Stimulat de problema determinării celei mai bune metode de selecţie a copiilor cu inteligenţă inferioară, Alfred Binet publică, în 1905, în "Année psychologique", în colaborare cu Théodore Simon, un articol intitulat *Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux*, în care prezintă 30 de probe pentru măsurarea inteligenţei în manifestările sale curente şi concrete. *Scala metrică a inteligenţei* elaborată de Binet şi Simon a avut un rol important în istoria testelor: ea este prima serie de teste mintale potrivită scopului pentru care a fost construită (clasificarea obiectivă a copiilor pe baza nivelului lor de inteligenţă, în funcţie de vârstă) şi a inspirat majoritatea realizărilor ulterioare în domeniul măsurării dezvoltării mintale. În 1916, Scala Binet-Simon a fost revizuită de către L.M. Terman, devenind cunoscuta *Scală Stanford-Binet*. L.M. Terman a sugerat multiplicarea coeficientului de inteligenţă cu 100 şi a propus utilizarea abrevierii "IQ" pentru coeficientul obţinut astfel.

Primii ani ai secolului al XX-lea reprezintă un moment de cotitură în teoria testelor psihologice.

Charles Spearman, în Anglia, elaborează metoda analizei factoriale şi, în 1914, publică lucrarea celebră *The Theory of two Factors*, în care afirmă structura bifactorială a inteligenţei: existenţa factorului G, al inteligenţei generale, şi a factorilor de grup. Această teorie a condus la elaborarea a numeroase teste mintale şi a stat la baza lucrărilor unor psihologi celebri, precum C. Burt, Ph. Vernon, G. Thomson şi L. L. Thurstone.

În aceeaşi perioadă, în Elveţia, C. G. Jung scrie primele lucrări relative la testul său de asociere a cuvintelor; acesta este prima tehnică proiectivă inspirată de concepţii freudiene. În 1920, Hermann Rorschach (1884-1922)

prezintă în "Psychodiagnostik" *Testul petelor de cerneală*, lucrare care constituie punctul de plecare al metodelor proiective. Dar, definiţia acestor metode nu va fi dată decât mai târziu, în 1939, de către Lawrence Frank, la patru ani după ce P. Rosenzweig şi-a prezentat testul proiectiv *Picture Frustration Study*, bazat pe teoria stresului şi a frustrării.

Un moment important în dezvoltarea metodei testelor îl reprezintă anul 1917, când se face prima administrare colectivă a testelor mintale. Este vorba despre instrumentele create de Robert Yerkes şi colaboratorii săi, *Army Alpha* şi *Army Beta*, destinate recrutării soldaţilor americani care urmau să fie trimişi să lupte în Europa. De la acea dată, testarea colectivă va fi utilizată pe scară largă, mai ales în orientarea şcolară şi în cercetările psihologice şi pedagogice.

Apariţia testelor de personalitate marchează o etapă nouă în dezvoltarea curentului psihometric. În 1914, Robert Woodworth a pus la punct primul chestionar de personalitate, *Personal Data Sheet*, în scopul de a standardiza interviurile psihiatrice. Câţiva ani mai târziu, Bruce Moore realizează un chestionar de interese, iar Edward Strong, un chestionar de interese vocaţionale. Utilizând analiza factorială, Raymond Cattell elaborează, în 1946, testul *Sixteen Personality Factors* (16PF), care are drept scop să evalueze şaisprezece trăsături fundamentale ale personalităţii. Ulterior, au apărut chestionare de personalitate, precum *Minnesota Multiphasic Personality Inventory* (MMPI) şi *California Psychological Inventory* (CPI), care investighează mai multe trăsături de personalitate, pentru ca, în perioada 1980-1990 să se accepte că numărul trăsăturilor fundamentale de personalitate poate fi redus la cinci şi să se pună bazele modului *Big Five*. Pornind de la această idee, Paul T. Costa şi Robert R. McCrae au pus la punct *Chestionarul de Personalitate NEO*, care evaluează cei cinci superfactori ai personalităţii: neuroticismul, extraversiunea, caracterul deschis, calitatea de a fi agreabil şi conştiinciozitatea. Ei sunt definiţi astfel (Aiken, 1994):

- neuroticismul: neliniştit vs. calm, nesigur vs. sigur, cu manifestări de autocompătimire vs. mulţumit de sine;
- extraversiunea: sociabil vs. retras, amator de distracţii vs. sobru, afectuos vs. rezervat;
- caracterul deschis: cu imaginaţie vs. ancorat în realitate; cu preferinţă pentru variaţie vs. cu preferinţă pentru rutină, independent vs. conformist;
- calitatea de a fi agreabil: sentimental vs. neîndurător; încrezător în alţii vs. suspicios, dornic să fie util vs. necooperant;
- conştiinciozitatea: organizat vs. dezorganizat, atent vs. neatent, autodisciplinat vs. lipsit de voinţă.

Faptul că în studiul personalităţii intervin mai ales elemente calitative ale comportamentului a făcut ca testul să nu mai poată fi privit ca o "măsură-

toare", cum se întâmpla în domeniul testelor de aptitudini. S-a ajuns astfel la o nouă definiţie a testului psihologic. Conform formulării date în 1933 de Asociaţia Internaţională de Psihotehnică, "testul este o probă definită, implicând o sarcină de realizat, identică pentru toţi subiecţii examinaţi, cu tehnici precise pentru aprecierea succesului ori a eşecului sau pentru notarea numerică a reuşitei". Considerând că această definiţie ar fi aplicabilă numai testelor de inteligenţă, de aptitudini şi de cunoştinţe, P. Pichot o propune, în 1967, pe următoarea: "Se numeşte test mintal o situaţie experimentală standardizată, care serveşte drept stimul unui comportament. Acest comportament este evaluat printr-o comparare statistică cu acela al altor indivizi plasaţi în aceeaşi situaţie, ceea ce permite clasarea subiectului examinat, fie cantitativ, fie tipologic" (Roşca, 1972).

După cel de-al doilea război mondial, utilizarea testelor s-a extins pe toate continentele, în domenii diverse: şcolar, profesional, terapeutic, social, militar, sportiv etc.

Teoria testelor a cunoscut completări şi modificări substanţiale. Teoria clasică, expusă de F. M. Lord şi M. R. Novick, în 1968, în lucrarea de referinţă *Statistical Theories of Mental Scores* a fost reformulată, într-o versiune modernă, de D. W. Zimmerman, în 1976. Dar, ea este treptat înlocuită de teoria răspunsului la itemi (Item Response Theory) şi de teoria stărilor şi a trăsăturilor latente (Latent State-Trait Theory).

Necesitatea de a cota şi interpreta rapid un număr mare de teste a condus, încă din 1930, la construirea unor maşini care să calculeze scorurile testelor. Douăzeci de ani mai târziu, s-a realizat în acest scop o maşină electronică dotată cu un dispozitiv fotoelectric. Acum, construirea, administrarea, cotarea şi interpretarea rezultatelor testelor psihologice beneficiază de ajutorul calculatoarelor electronice. Datorită lor se poate pune în aplicare o idee mai veche, cea de "adaptare" a testului la particularităţile subiectului, astfel încât, utilizându-se cât mai puţini itemi să se obţină cât mai multă informaţie despre persoana examinată. Este vorba de aşa-numitele *teste psihologice adaptative.*

3.2. CÂTEVA TEORII ALE PERSONALITĂŢII ŞI IMPACTUL LOR ASUPRA PSIHOMETRIEI

3.2.1. Definiţii

A. În contextul teoriilor referitoare la personalitate, **situaţia** este înţeleasă ca un ansamblu format din patterne de stimuli, evenimente, persoane

şi factori afectivi care există la un moment dat (Reber, 1985). Conform dicţionarului editat de Petrovsky şi Yaroshevsky (1985), o situaţie este un sistem de condiţii care induc şi mediază activitatea unei persoane, dar este exterior acesteia în ceea ce priveşte :

- spaţiul (este localizat în exteriorul subiectului);
- timpul (este anterior acţiunilor subiectului);
- funcţionarea (acţionează independent de subiect).

În cazul testării psihologice, printre elementele situaţiei se află informaţiile comunicate subiecţilor despre natura şi scopul testării, condiţiile de administrare a testului, starea subiectului din momentul precedent testării -dacă aceasta are influenţă asupra comportamentului său-, precum şi consecinţele pe care le-au avut asupra subiectului performanţele realizate la alte testări.

B. Termenul trăsătură a fost definit în diverse moduri. El a fost utilizat în acelaşi sens ca şi "dimensiune", "dispoziţie" sau "caracteristică stabilă".

Într-una din accepţiuni, o **trăsătură** este o etichetă atribuită unui ansamblu de comportamente observabile şi stabile, care îi deosebesc pe indivizi între ei. De exemplu, J. P. Guilford priveşte trăsătura ca fiind "o manieră durabilă în care o persoană diferă de altele" (Corsini, 1994). În altă accepţiune, trăsătura este o "dimensiune" de-a lungul căreia indivizii ocupă poziţii diferite, în funcţie de "tendinţele pe care le au de a produce patterne consistente de gândire, simţire şi acţiune" (McCrae, Costa, 1995). Ocupându-se de raportul dintre trăsăturile de personalitate ca variabile dispoziţionale care fixează structura personalităţii şi determinanţii situaţionali, M. Golu afirmă că "o trăsătură este privită ca o mulţime de comportamente care tind să covarieze (să fie corelate) într-o manieră relativ constantă în decursul timpului, în grupe de indivizi şi în variate situaţii, în care un individ se deosebeşte de alţii" (Marcus, 1997).

După unii autori, trăsăturile sunt constructe de personalitate create de psihologi, deci noţiuni abstracte, fără existenţă reală, care se manifestă prin diverşi "indicatori ai trăsăturii" (patterne de motive, atitudini, comportamente) (McCrae, Costa, 1995). După alţii, trăsăturile sunt caracteristici, calităţi sau procese care există în persoane, fiind, deci, o realitate obiectivă. De exemplu, G. W. Allport consideră că trăsătura este "o structură neuropsihică". În accepţiunea sa, o trăsătură îi predispune pe indivizi să perceapă stimuli diferiţi ca având aceeaşi semnificaţie şi să răspundă la astfel de stimuli prin comportamente similare (Hall, Lindzey, 1985).

C. În psihologia ultimelor decenii s-a manifestat tendinţa de a face distincţie între *stare* (înţeleasă ca reacţie trecătoare la o situaţie) şi *trăsătură* (privită ca o caracteristică relativ stabilă a personalităţii).

Lucrările elaborate de C. D. Spielberger şi colaboratorii săi enumeră următoarele caracteristici ale stărilor şi trăsăturilor (Spielberger, Gorsuch, Lushene, 1970; Spielberger, 1988; Steyer, Ferring, Schmitt, 1992):

O *stare* (a personalităţii) este prezentă la un anumit moment. Intensitatea sa poate fi mai mare sau mai mică, în funcţie de situaţia care a generat-o. Deşi stările sunt adesea trecătoare, ele pot să reapară atunci când sunt provocate de stimuli potriviţi şi pot să dureze atât timp cât condiţiile care le-au determinat persistă. Reacţiile emoţionale sunt expresii ale stărilor de personalitate.

Trăsătura este conceptualizată ca o tendinţă relativ stabilă a individului de a percepe lumea într-un anumit mod şi/sau de a reacţiona ori a se comporta într-o anumită manieră, cu o regularitate predictibilă. Intensitatea unei trăsături are influenţă asupra frecvenţei de manifestare şi a intensităţii stării corespunzătoare. Mai precis, cu cât o trăsătură este mai accentuată la o persoană, cu atât este mai mare probabilitatea ca aceasta să prezinte starea emoţională care corespunde trăsăturii în situaţii mai variate şi cu o intensitate mai mare.

De exemplu, supărarea a fost definită ca stare şi ca trăsătură astfel:

• Starea supărare este o stare emoţională care se manifestă prin încordare, iritare şi furie, însoţite de tensiune musculară şi de o activare a sistemului nervos vegetativ. Ea este îndreptată de obicei împotriva prietenilor şi a persoanelor iubite şi doar rareori asupra străinilor sau a persoanelor pe care subiectul nu le simte apropiate (Averill, 1980).

• Trăsătura supărare reprezintă dispoziţia subiectului de a aprecia o gamă largă de situaţii ca fiind enervante sau nemulţumitoare şi tendinţa sa de a reacţiona frecvent la acestea prin creşterea intensităţii stării de supărare (Spielberger, 1988).

Au existat numeroase încercări de a identifica atributele care diferenţiază stările de trăsături. Dar, nu a putut fi formulată nici o condiţie necesară sau suficientă pentru a recunoaşte o stare sau o trăsătură. S-a recurs atunci la definirea categoriilor "stare" şi "trăsătură" prin câte un grup de atribute care corelează între ele (Chaplin, John, Goldberg, 1988; John, 1990). Astfel:

• Stările sunt temporare, inconsistente de la o situaţie la alta, sunt de scurtă durată şi au cauză externă. Stările identifică acele comportamente care pot fi controlate prin manipularea situaţiei.

• Trăsăturile sunt stabile în timp, consistente de la o situaţie la alta, sunt de durată lungă, au cauză internă şi pentru ca prezenţa lor să fie constatată este necesar să fie efectuate mai multe observaţii, în situaţii variate, decât în cazul stărilor. Trăsăturile permit predicţia viitorului pe baza trecutului.

3.2.2. Câteva teorii ale personalităţii

Deosebirea esenţială dintre diversele teorii referitoare la personalitate constă în ponderea pe care o acordă rolului jucat de factorii interni (legaţi de *persoană*) şi de cei externi (legaţi de *situaţie*) în declanşarea, menţinerea şi modificarea comportamentului. Căutând să răspundă la întrebarea "de ce variază comportamentul intrapersonal şi cel interpersonal?", adepţii acestor teorii au efectuat cercetări prin care au încercat să determine "când apar variaţii ale comportamentului". Fiecare teorie şi-a constituit o metodologie proprie de cercetare, cu un mod specific de culegere şi prelucrare a datelor.

Multă vreme, până prin anii 1960-1970, teoriile şi cercetările referitoare la personalitate au fost dominate de **modelul trăsăturilor** (trait model). Conform acestuia, comportamentul (C) este determinat de predispoziţii latente stabile ale persoanei (P), numite **trăsături**. Adică:

C = f(P).

Modelul trăsăturilor oferă posibilitatea de a concentra într-un set limitat de trăsături caracteristici afective, atitudinale şi comportamentale ale indivizilor. Dar, fiind o teorie descriptivă, el nu identifică mecanismele care stau la baza trăsăturilor. În acest model, "un psihograf al unei trăsături este ca un buletin meteorologic, bun pentru a vă spune dacă trebuie să purtaţi umbrela, dar sărac în furnizarea de explicaţii de ce ar putea să plouă" (Ozer, Reise, 1994).

În decursul timpului au fost elaborate **modele cu o singură trăsătură** (single-trait models) şi **modele cu mai multe trăsături** (multi-traits models).

Pentru modelele din prima categorie, trăsătura este un aspect specific al personalităţii, care îndeplineşte următoarele condiţii:

- este prezent la toate persoanele, dar în grade diferite;
- este invariabil intraindividual;
- joacă un rol important în determinarea comportamentului;
- poate fi măsurat cu precizie.

Aceste modele utilizează instrumente (scale, teste, chestionare) prin care evaluează câte o singură trăsătură. În lucrările referitoare la personalitate, despre o asemenea trăsătură se spune că este "de nivel inferior" sau "specifică". Specificitatea se referă la contextul în care este definită trăsătura, respectiv la varietatea situaţiilor în care trebuiesc observate comportamentele care servesc la definiţia trăsăturii respective. Cu cât numărul acestor situaţii este mai mic şi ele sunt mai asemănătoare între ele, cu atât trăsătura are un grad de specificitate mai mare. Spre exemplu, se vorbeşte despre trăsătura "anxietate", dar şi despre trăsăturile "anxietate faţă de calculator" şi "anxietate faţă de teste". În primul caz trăsătura are o generalitate mai mare, o specificitate mai mică. Precizia cu care poate fi prezis un anumit comportament într-o situaţie particulară (cum ar fi

lucrul la calculator sau examenul psihologic) este mai mică pentru trăsăturile generale decât pentru trăsăturile specifice situaţiei.

Unii autori au definit trăsături "legate de sarcină" (task-related personality traits). Este vorba despre trăsături, mai mult sau mai puţin generale, care pot servi la predicţia succesului într-un anumit domeniu de activitate. Aşa sunt, de exemplu, trăsăturile educaţionale studiate de B. de Raad (1996).

Modelele cu mai multe trăsături operează la un nivel de abstractizare mai înalt decât modelele cu o singură trăsătură. Ele presupun că există un număr redus de trăsături fundamentale, globale, numite uneori "factori de nivel înalt" sau "superfactori", pe care încearcă să le identifice prelucrând prin analiză factorială rezultatele unor chestionare, observaţii şi teste obiective referitoare la trăsături specifice. Factorii rezultaţi sunt consideraţi ca surse ale trăsăturilor specifice (Boekaerts, 1996). Cele mai importante modele cu mai multe trăsături sunt: modelul lui Eysenck bazat pe neuroticism, extraversiune şi psihoticism, teoria tipurilor de personalitate şi modelul Big Five.

În mod firesc, apare întrebarea: care trăsături este mai important să fie evaluate de către psihologi, cele specifice sau cele globale? Cercetătorii susţin că alegerea trebuie dictată de scopul examinării psihologice: pentru "explicare", sunt mai utile trăsăturile globale, în timp ce pentru "predicţie", cele specifice (Lay, 1997).

În momentul definirii ei, fiecare trăsătură este pusă în legătură cu un grup de comportamente, care servesc ca indicatori ai acesteia. De exemplu, J.P. Guilford defineşte empatia ca "o abilitate de a prezice, de a recunoaşte dispoziţiile psihologice ale unei alte persoane, percepţiile sale, gândurile, sentimentele şi atitudinile, de asemenea, ca trăsături proprii" (Marcus, 1997). Pentru aprecierea gradului în care trăsătura este prezentă la fiecare persoană, se utilizează un model aditiv, care constă din următorii paşi:

(1) Se evaluează unul sau mai multe din următoarelor criterii (G. W. Allport, după Corsini, 1994):

- frecvenţa cu care apar comportamentele;
- varietatea situaţiilor în care apar;
- intensitatea cu care se manifestă;
- durata lor.

(2) Se cumulează măsurile obţinute.

(3) Pe baza valorii rezultate, se fac inferenţe asupra măsurii trăsăturii. În acest scop se aplică un algoritm specific instrumentului utilizat la evaluarea criteriilor sau se compară valoarea obţinută pentru o persoană cu cele ale altor indivizi.

Testele şi chestionarele folosite cel mai des pentru colectarea datelor, atât în cercetare cât şi în psihologia aplicată (de exemplu, în clasificare, selecţie,

consultanţă) au la bază modelul trăsăturilor. De asemenea, multe metode matematice utilizate frecvent de psihologi pornesc de la ipoteza existenţei unor variabile latente (corespunzătoare trăsăturilor) care sunt stabile în timp şi se manifestă prin intermediul variabilelor observate. Aşa sunt: analiza factorială, teoria clasică a testelor şi teoria răspunsului la itemi.

Analiza factorială reduce o mulţime de variabile observate (scoruri de itemi, scale sau teste, rezultatele unor observaţii etc.), corelate liniar între ele, la un număr mai mic de variabile necorelate (factorii extraşi), cu ajutorul cărora se poate realiza o descriere obiectivă a persoanelor. Un **factor** este o combinaţie liniară ponderată a variabilelor prelucrate prin analiză factorială. Ponderile rezultă din calcule şi sunt determinate de valorile coeficienţilor de corelaţie liniară dintre variabilele măsurate. Factorii extraşi sunt concepte statistice. Ei reprezintă variabile latente (ce nu pot fi observate direct), care au legătură de corelaţie liniară cu variabilele observate prelucrate prin analiza factorială. Ei nu corespund totdeauna unor caracteristici reale ale persoanelor sau unor trăsături. Semnificaţia fiecărui factor se deduce în urma analizei valorilor ponderilor variabilelor observate.

Teoria clasică a testelor foloseşte descompunerea *scorului observat* la test, scală sau item în *scor real* şi în *eroare de măsură.* Scorul observat reprezintă rezultatul unei măsurători, în timp ce scorul real corespunde unei variabile latente. Dacă instrumentul utilizat în investigaţie serveşte la evaluarea unei trăsături, atunci scorul real este tocmai măsura trăsăturii respective. Teoria clasică a testelor utilizează în studiul relaţiei dintre variabile modelul liniar. În consecinţă, gradul de acord dintre variabile este exprimat prin intermediul unui coeficient de corelaţie liniară.

În teoria răspunsului la itemi, spre deosebire de teoria clasică a testelor, relaţia dintre scorurile itemilor şi valoarea variabilei latente măsurate de test este exprimată analitic printr-o funcţie de regresie neliniară. Dacă itemul funcţionează corect, funcţia de regresie este monotonă (crescătoare sau descrescătoare).

Teoriile referitoare la personalitate au analizat variabilitatea comportamentelor sub două aspecte: al *consistenţei* (constanţa comportamentului în diverse situaţii existente la momente foarte apropiate între ele) şi al *stabilităţii temporare* (constanţa comportamentului în situaţii asemănătoare dar distanţate în timp). Cercetările efectuate au evidenţiat grupe de comportamente asemănătoare în ceea ce priveşte variabilitatea intrapersonală şi categorii de persoane pentru care unele comportamente se manifestă asemănător în privinţa variabilităţii.

Astfel, s-a observat că atât consistenţa cât şi stabilitatea temporară sunt mai mari pentru comportamentele care sunt asociate cu inteligenţa sau cu

procese cognitive. Etichetele referitoare la personalitate, pe care le folosesc indivizii pentru a se descrie în scalele de autoevaluare, au o stabilitate temporară mare (Mischel, 1968). E. L. Kelly, administrând mai multe chestionare unor cupluri, de două ori, la un interval de 16-19 ani, a constatat că, spre deosebire de atitudinile faţă de căsătorie, care s-au schimbat mult şi în diverse moduri, încrederea în sine, sociabilitatea, unele interese, unele valori economice şi politice evaluate prin chestionare au avut o stabilitate temporară mare (Mischel, 1968).

Cercetările experimentale au condus la două observaţii importante:

• Persoanele se deosebesc între ele în ceea ce priveşte variabilitatea comportamentelor de la o ocazie (situaţie sau moment de observare) la alta, unele fiind mai schimbătoare, iar altele mai consecvente. Unii autori susţin că pentru fiecare aspect al personalităţii există trăsături generale care au rol moderator asupra variabilităţii intrapersonale. De exemplu, R. J. Larsen şi colaboratorii săi au observat că variaţia ciclică săptămânală a dispoziţiilor este mai evidentă la introvertiţi decât la extravertiţi (Penner, Shiffman ş.a., 1994).

• La un acelaşi individ, consistenţa comportamentală de la o situaţie la alta se manifestă asemănător pentru mai multe comportamente, nu doar pentru cele specifice unei anumite trăsături (Amelang, Borkenau, 1986). Pornind de la această observaţie, s-a definit o variabilă continuă, considerată de unii ca fiind o "trăsătură", care reprezintă consistenţa comportamentală generală. Persoanele cu valori mari ale acestei variabile sunt doar puţin dependente de situaţie.

S-a încercat să se determine cauza care face ca la o persoană comportamentele asociate cu unele trăsături să varieze mult de la o situaţie la alta, iar cele legate de alte trăsături, să fie constante. S-a observat că trăsăturile cărora le corespund comportamente puţin variabile sunt cele pe care indivizii le consideră "relevante" pentru ei (adică potrivite pentru a-i descrie) sau "observabile" (adică destul de evidenţiate pentru a putea fi observate de alţii) (Amelang, Borkenau, 1986). Probabil că, într-un profil al personalităţii care ar îngloba foarte multe trăsături, acestora le-ar corespunde punctele de extrem (de minim sau de maxim).

O posibilitate de evaluare a constanţei (consistenţei şi stabilităţii temporare) unei trăsături pentru o persoană constă în analizarea similarităţii profilurilor corespunzătoare mai multor manifestări comportamentale, construite în situaţii diferite, şi a similarităţii profilurilor corespunzătoare unor situaţii diverse, construite pentru diferite comportamente care au legătură cu trăsătura respectivă. Similaritatea perfectă este considerată o dovadă a constanţei individuale (Schmitt, Borkenau, 1992).

Pe o poziţie opusă modelului trăsăturilor se situează **modelul situaţionist** (situationism model), ale cărui baze au apărut pe la începutul seco-

lului al XX-lea. El a fost puternic influenţat de sociologie, datorită numărului mare de cercetări efectuate asupra rolului mediului (al factorilor geografici, socio-economici, profesionali, familiali etc.) în dezvoltarea personalităţii. Conform acestui model, rolul principal în determinarea comportamentului (C) îl au factorii situaţionali sau stimulii prezenţi în fiecare situaţie (S):

C = f(S)

Modelul situaţionist neagă rolul caracteristicilor individuale, al factorilor intrapsihici şi al antecedentelor personale, asupra comportamentului. Adepţii acestui model au considerat că relaţia persoană-mediu poate fi reprodusă în laborator prin diverse experimente, iar variabilele independente şi cele dependente pot fi clar identificate şi operaţionalizate în factori cauzali şi efecte.

Reprezentanţii situaţionismului nu au creat metode noi de colectare sau de prelucrare a datelor. În general culegerea datelor s-a bazat pe contabilizarea răspunsurilor comportamentale, iar metoda statistică aplicată cel mai frecvent a constituit-o analiza de varianţă.

Deşi bazele sale teoretice au fost puse în perioada 1920-1930, **modelul interacţionist** (interactionism model) s-a dezvoltat doar după anul 1960. Prima conferinţă internaţională asupra interacţiunii om-situaţie a avut loc la Stockholm în 1975.

Modelul interacţionist recunoaşte influenţa mediului asupra comportamentului, dar, în acelaşi timp, evidenţiază rolul factorilor personali în determinarea acţiunii. Conform acestui model, comportamentul este rezultatul unei interacţiuni continue între mecanismele psihologice şi factorii de mediu:

C = f(P, S)

Teoriile interacţioniste au identificat următoarele moduri de interacţiune între persoană şi situaţie (Hettema, Kenrick, 1992):

a. interacţiune statică: atât trăsăturile persoanelor cât şi caracteristicile mediului rămân relativ neschimbate o perioadă mai îndelungată;

b. persoana alege situaţia: deşi uneori confruntarea persoanelor cu situaţiile este întâmplătoare, de obicei persoanele caută mediul care convine cel mai mult caracteristicilor proprii şi din acest motiv ocolesc sau părăsesc situaţiile care nu le satisfac;

c. situaţia alege persoana: unele situaţii sunt accesibile numai anumitor persoane (de exemplu, nu orice absolvent de liceu poate să îşi continue studiile la o facultate);

d. persoana transformă situaţia: se presupune că persoana şi-a formulat un scop şi acţionează pentru a-l realiza, modificând situaţia;

e. situaţia transformă persoana: persoana îşi adaptează comportamentul la cerinţele situaţiei, iar dacă este confruntată repetat cu aceleaşi constrângeri, îşi modifică propriile caracteristici;

f. persoana şi situaţia tranzacţionează între ele: persoana schimbă unele caracteristici ale mediului pentru a-l face convenabil trăsăturilor sale, dar în acelaşi timp îşi schimbă unele caracteristici (comportamente, acţiuni, scopuri, gânduri, sentimente, motive), pentru a se adapta situaţiei.

Interacţionismul clasic punea accentul pe interacţiunea reciprocă, dinamică şi continuă, dintre individ şi mediu. Modelele moderne ale interacţionismului se ocupă cu interacţiunea dintre acţiuni şi scopuri, dintre emoţie şi cogniţie, dintre factorii biologici şi cei sociali, acordând o importanţă deosebită studiului adaptării şi al copingului la stres.

Adepţii modelului interacţionist au considerat că metodele utilizate până atunci pentru culegerea şi prelucrarea datelor nu sunt potrivite pentru cercetarea personalităţii. Ei au înlocuit chestionarele obişnuite cu chestionarele de tip S-R (situaţie-răspuns), dezvoltate de N. S. Endler, J. McV. Hunt şi A. J. Rosenstein. În aceste chestionare răspunsurile şi situaţiile sunt luate în considerare simultan.

La prelucrarea datelor, modelul interacţionist utilizează metode noi, precum modelul Markov pentru studierea proceselor în care intervin interacţiuni continue şi modelele descriptive pentru formarea grupelor omogene de indivizi pe baza patternurilor de reacţii la situaţii diverse (Endler, Magnusson, 1976). **Teoria stărilor şi a trăsăturilor latente,** dezvoltată în ultimul deceniu, are la bază modelul interacţionist.

3.3. EVALUARE PSIHOMETRICĂ SAU EVALUARE COMPORTAMENTALĂ ?

Evaluarea este un domeniu al psihologiei ştiinţifice -în parte metodologică, în parte aplicativă- care se ocupă cu explorarea, aprecierea şi măsurarea unor evenimente psihologice (Fernández-Ballesteros, 1993)

Evaluarea psihologică are drept scop să aprecieze individul, în general în legătură cu una sau mai multe probleme specifice, cum sunt: funcţiunile intelectuale, trăsăturile de personalitate, aptitudinile speciale, problemele de sănătate, dificultăţile de învăţare, comportarea şcolară, diverse variabile emoţionale şi sociale. Ea a apărut o dată cu testele de inteligenţă, la începutul acestui secol. S-a dezvoltat pe măsură ce au fost elaborate testele de personalitate, educaţionale şi vocaţionale, din necesitatea de a se pune la punct proceduri statistice care să permită crearea, analizarea şi îmbunătăţirea testelor. Teoria şi tehnologia instrumentelor propuse în acest scop au fost denumite "psihometrice", iar evaluarea bazată pe ele, **evaluare psihometrică**. Ea reprezintă una dintre variantele evaluării psihologice tradiţionale.

Faptul că apariţia psihometriei este legată de utilizarea pe scară largă a testelor a făcut ca denumirile de "metoda testelor" şi "metoda psihometrică" să fie uneori folosite ca sinonime. În principal, psihometria se ocupă cu analiza de itemi, cu studiul validităţii şi al fidelităţii şi cu etalonarea testelor.

Prin anii 1960 în evaluarea psihologică a apărut un curent nou: **evaluarea comportamentală.** Aceasta are ca obiectiv măsurarea directă a comportamentului subiecţilor. Ea încearcă să descopere *ce* fac indivizii, *unde*, *când* şi *în ce circumstanţe*. Baza pentru evaluarea şi predicţia comportamentului o formează, în această abordare, comportamentul motoriu şi verbal manifest; nu se mai fac inferenţe sau interpretări ale unor rezultate de test.

Expresia "evaluare comportamentală" a devenit comună în cercetările legate de modificările de comportament şi în terapie la jumătatea deceniului al optulea, când se considera că abordarea comportamentală are rolul "de a identifica şi măsura unităţile de răspuns şi variabilele care le controlează (ce ţin de mediu sau de organism), în scopul înţelegerii şi al modificării comportamentului uman" (R. O. Nelson şi S. C. Hayes, citat după Silva, 1993).

De la apariţia sa, evaluarea comportamentală s-a delimitat de evaluarea psihometrică, atât în privinţa scopului, cât şi a metodelor utilizate. Sunt foarte concludente în acest sens normele metodologice pe care şi le-a fixat (Fernández-Ballesteros, 1993):

- Principala metodă de măsurare este observarea comportamentului.
- Chestionarele de personalitate sunt incapabile să descrie probabilitatea răspunsurilor sau variabilele care afectează aceste probabilităţi, deci ele vor fi abandonate.
- Se va renunţa la testele tradiţionale, care evaluează trăsături psihice (cum sunt structurile mintale), şi la principiile lor psihometrice.

Dar, deşi adepţii evaluării comportamentale au opus rezistenţă psihometriei, aceste reguli nu au putut fi respectate. Din contră, în decursul ultimelor două decenii au existat mai multe tentative de apropiere a evaluării comportamentale de evaluarea psihometrică. Astfel, s-au construit instrumente de evaluare a comportamentului utilizându-se tehnici statistice, s-au făcut studii asupra structurii factoriale a chestionarelor, s-a cercetat validitatea concurentă a măsurătorilor furnizate de observaţii sau de interviuri. În repetate rânduri s-a subliniat necesitatea de a se lucra şi în evaluarea comportamentului cu instrumente corespunzătoare ca validitate şi fidelitate, care să fie verificate prin metode psihometrice.

Constatarea că există factori care nu pot fi măsuraţi direct, precum motivaţia, dar care influenţează comportamentul a avut drept consecinţă restrângerea utilizării procedurilor de observare a comportamentului şi utilizarea din ce în ce mai frecventă a chestionarelor de autoevaluare.

Testele psihologice continuă să fie aplicate în practică şi de susţinătorii evaluării comportamentale. De pildă, în revistele *Behavioral Assessment* (în perioada 1980-1987) şi *Journal of (Psychopathology and) Behavioral Assessment* (în perioada 1979-1986) mai mult de 25% dintre articole se ocupau cu construcţia de teste (Fernández-Ballesteros, 1993). Dar, se poate observa tendinţa ca fiecare tip de evaluare -psihometrică şi comportamentală- să folosească instrumente specifice de măsurare. Fernández-Ballesteros (1993) a prelucrat, prin analiza corespondenţelor, articolele publicate în perioada 1989-1991, în revistele *Behavioral Assessment, Journal of Psychopathology and Behavioral Assessment* şi *Psychological Assessment*, din punctul de vedere al metodologiilor aplicate şi a constatat formarea a două grupări separate net şi aflate la poluri opuse ale aceleiaşi axe: una conţine lucrările despre construcţia de teste, tehnicile proiective, scalele psihopatologice, testele de inteligenţă şi MMPI (apărute, evident, în *Psychological Assessment*), cealaltă cuprinde articolele referitoare la observaţii, design experimental, chestionare de autoevaluare şi interviuri (publicate în *Behavioral Assessment*).

De-a lungul timpului, evaluarea comportamentală şi-a sporit interesul pentru metodologie. Au fost create metode proprii de măsurare (de exemplu, tehnici de observare, proceduri fiziologice, interviuri structurate, tipuri noi de scale de evaluare a persoanelor de către alte persoane etc.).

Analizând toate rezumatele articolelor din *PsycLIT* (*Psychological Abstract* în versiunea CD-ROM) referitoare la evaluarea comportamentală, apărute în anii 1980-1982 şi 1990-1992, Cone (1993) observă între cele două perioade o creştere a frecvenţei lucrărilor cu conţinut metodologic (de la 31.8% la 46.0%) şi o scădere a frecvenţei celor cu conţinut aplicativ (de la 40% la 25.5%).

În cadrul evaluării comportamentale au fost elaborate şi utilizate, în ultimii zece ani, un număr foarte mare de modele, strategii, instrumente şi practici de evaluare. Această observaţie l-a condus pe D. J. Cone, în 1986, la concluzia că "nu există o singură evaluare comportamentală, ci mai multe" (Silva, 1993). Totuşi, în 1988, D. J. Cone reduce toate modelele de evaluare comportamentală la două, opuse din punctul de vedere al abordării pe care o au la bază (Silva, 1993):

- **abordarea nomotetică a trăsăturii** (nomothetic-trait approach), care are ca obiect de studiu trăsături sau sindroame (de exemplu, anxietatea socială, depresia, agorafobia), dezvoltă instrumente pentru evaluarea lor deductivă şi stabileşte adecvanţa acestor instrumente pe baza modului în care variază scorurile lor între indivizi; instrumentele sunt folosite pentru a măsura efectele variabilelor independente în cadrul grupelor de subiecţi sau pentru a evalua efectele intervenţiilor aplicate indivizilor;

- **abordarea idiografică a comportamentului** (idiographic-behavior approach), care are ca obiect de studiu un comportament specific, dezvoltă procedee pentru evaluarea sa inductivă şi stabileşte adecvanţa acestor proceduri ţinând cont de modul în care variază scorurile aceluiaşi individ.

Abordarea nomotetică are la bază presupunerile că atunci când instrumentele de evaluare sunt precise se pune în evidenţă o consistenţă a comportamentului persoanelor şi că măsurile trăsăturilor permit o predicţie bună a comportamentului (Pushton, Jackson, Paunonen, 1981). Ambele presupuneri au fost infirmate de practică. Observaţia că puterea predictivă a testelor de personalitate pentru un anumit comportament este, în general, destul de scăzută, a fost explicată prin faptul că abordarea nomotetică se bazează pe "ideea falsă" că orice trăsătură este "universal aplicabilă tuturor persoanelor", când, în realitate, fiecare individ are anumite trăsături relevante. Cele care nu îi sunt relevante nu permit predicţia comportamentului care este caracteristic acelei trăsături (Paunonen, Jackson, 1985). Opusă acestei concepţii, abordarea idiografică susţine că fiecărei persoane îi sunt specifice anumite determinante ale comportamentului şi pe acestea încearcă să le identifice. În abordarea nomotetică se caută legile generale sau ecuaţiile care se aplică tuturor persoanelor, în timp ce în abordarea idiografică se încearcă să se determine ecuaţiile proprii fiecărui individ.

Procedura statistică la care face apel abordarea nomotetică este analiza de regresie, mai precis construirea funcţiei de regresie care leagă predictorul (o trăsătură de personalitate) de criteriu (un comportament). Pentru obţinerea ecuaţiei funcţiei de regresie se utilizează date de la un eşantion, iar ecuaţia obţinută serveşte la predicţia comportamentului tuturor indivizilor din populaţie. Prin strategiile idiografice se identifică persoanele pentru care este potrivită o anumită ecuaţie de predicţie a comportamentului pe baza trăsăturii.

Se poate vedea că termenul "abordare nomotetică a trăsăturii" defineşte de fapt concepţia care stă la baza evaluării tradiţionale (Silva, 1993). Unii psihologi exclud acest tip de abordare din evaluarea comportamentală.

Încercările de îndepărtare a evaluării comportamentale de evaluarea psiholometrică pun accentul pe deosebirile existente între ele. Una dintre acestea se referă la presupunerile făcute asupra comportamentului.

În teoria psihometrică, scorul observat la un test este compus dintr-un scor real şi o eroare de măsură. Scorul real reflectă o entitate internă stabilă, deci nu se schimbă (cel puţin într-un interval scurt de timp). Din acest motiv, un instrument de măsură este apreciat ca fiind bun, dacă scorurile observate la o aceeaşi persoană, în două ocazii, nu diferă prea mult.

În concepţia evaluării comportamentale, comportamentul nu este, în mod necesar, stabil. Variaţia scorurilor observate la un test, pentru un acelaşi individ, poate fi cauzată de schimbările produse în comportamentul acestuia şi nu neapărat de imprecizia intrumentului de măsură. Atunci când apelează, totuşi, la teste, evaluarea comportamentală caută să explice cauzele schimbării, să determine variabilele responsabile de aceasta şi nu face nici o presupunere asupra erorii de măsură.

Întrucât evaluarea comportamentală susţine dependenţa de situaţie a comportamentului, ea nu utilizează validarea concurentă ca mijloc de predicţie şi nici nu pretinde să se obţină valori mari ale coeficientului de fidelitate test-retest.

Cu toate deosebirile existente între ele, evaluarea psihometrică şi cea comportamentală prezintă acum mai multe similarităţi decât în urmă cu trei decenii (Fernández-Ballesteros, 1993).

3.4. TENDINŢE NOI ÎN STUDIUL PERSONALITĂŢII: MODELUL BIG FIVE ŞI MODELUL CIRCUMPLEX

3.4.1. Modelul Big Five

Modelul Big Five este un produs al abordării lexicale a personalităţii. La baza acestei abordări stau următoarele două ipoteze fundamentale (Saucier, Goldberg, 1996):

- Cele mai răspândite, semnificative şi distinctive atribute fenotipice tind să fie exprimate printr-un singur cuvânt.
- Gradul de reprezentare a unui atribut în limbaj tinde să corespundă importanţei relative a atributului.

Altfel spus, "acele diferenţe individuale care au cea mai mare semnificaţie în tranzacţiile zilnice dintre persoane vor fi, în cele din urmă, encodate în limbajul lor. Cu cât este mai importantă o asemenea diferenţă, cu atât mai mult indivizii o vor remarca şi vor vorbi despre ea, rezultatul fiind că vor ajunge să inventeze un cuvânt pentru ea (L.R. Goldberg, după Hendriks, 1997).

Pornind de la aceste presupuneri, o modalitate de a descrie domeniul trăsăturilor o constituie realizarea unui dicţionar redus, care să conţină toţi termenii relevanţi referitori la personalitate. Primul care a făcut acest lucru se pare că a fost Sir Francis Galton, în 1884. Dar, mai importantă decât lista sa este cea extrasă de G. W. Allport şi H. S. Odbert din *Webster's New International*

Dictionary (ediţia din 1925), formată din aproape 18000 de descriptori, cuvinte care deosebesc comportamentul unei persoane de cel al alteia. Dintre aceşti termeni, 4504 se referă la trăsături. Se poate afirma că ei stau la originea modelului cu cinci factori.

În 1934, L. L. Thurstone a apelat la 1300 de persoane pentru a evalua, fiecare, un individ pe care îl cunoaşte bine, folosind o listă compusă din 60 de adjective. Prelucrând prin analiză factorială rezultatele obţinute, el a ajuns la concluzia că personalitatea poate fi explicată prin cinci factori independenţi.

Pe de altă parte, R. B. Cattell, pornind de la lista cu cei 4504 termeni referitori la trăsături a lui Allport şi Odbert, a aplicat analiza factorială şi, prin rotaţie oblică, a obţinut 12 factori, dintre care primii cinci apar cu saturaţii mari. Ei sunt greu de interpretat, dar Cattell susţine că aceştia reprezintă cele mai importante dimensiuni ale personalităţii (John, 1990).

Adevăraţii descoperitori ai modelului Big Five sunt, însă, E. C. Tupes şi R. E. Christal, care au analizat mai multe lucrări dedicate scalelor bipolare construite de Cattell şi au găsit cinci factori de personalitate.

În anii următori, numeroase studii efectuate asupra unor chestionare de evaluare sau de autoevaluare şi asupra unor liste de adjective ori de verbe, realizate pe eşantioane de subiecţi de sexe, vârste şi naţionalităţi diferite şi prin metode diferite de analiză au pus în evidenţă câte cinci factori, în majoritatea cazurilor asemănători între ei. Ei au fost denumiţi de Lewis Goldberg, în 1981, *Big Five*, pentru că fiecare dintre ei acoperă un domeniu foarte larg şi cuprinde un număr mare de caracteristici de personalitate, distincte între ele şi specifice fiecărui factor.

Deşi numele atribuit factorilor diferă de la un autor la altul, consensul în ceea ce priveşte conţinutul psihologic al acestora este destul de mare. Prezentăm în continuare semnificaţia factorilor şi câteva dintre denumirile sub care sunt cunoscuţi (John, 1990; Zlate, 1994; Minulescu, 1996):

- Extraversia, sociabilitatea şi ambiţia, atracţia interpersonală, afirmarea, puterea, izbucnirea (surgency) - arată capacitatea de orientare a personalităţii către exterior, modul de implicare în acţiune, sociabilitatea persoanei.

- Agreabilitatea, adaptabilitatea socială, plăcerea, nivelul socializării, amabilitatea, dragostea - se referă la calităţile emoţionale ale persoanei şi la comportamentele ei prosociale.

- Conştiinciozitatea, conformitatea, dependenţa, responsabilitatea, prudenţa, autocontrolul, interesul pentru muncă, puterea realizării - vizează modul concret, caracteristic al individului de a trata sarcinile, activităţile, problemele care apar în viaţa lui; cuprinde trăsături cum ar fi ordinea, disciplina, responsabilitatea socială.

• Stabilitatea emoţională, controlul emoţional, emoţionalitatea, neuroticismul, adaptarea, anxietatea, afectul - arată caracteristicile emoţionale ale persoanei (calm, mulţumit, emotiv) şi diferite dificultăţi emoţionale (anxietate, depresie, iritabilitate).

• Cultura, intelectul, curiozitatea intelectului, inteligenţa, subtilitatea, deschiderea, independenţa, grija/atenţia - se referă la funcţiile intelectuale (creativitate, inventivitate, deschidere la experienţă).

Deosebirile existente între conţinuturile factorilor şi, în consecinţă, între denumirile atribuite lor, la diverşi autori, provin din faptul că elementele introduse în prelucrare prin analiză factorială (cuvinte care exprimă trăsături, itemi, scale etc.) nu sunt indicatori puri ai factorilor (J. A. Johnson, F. Ostendorf, după Johnson, 1996). De obicei, fiecare element are o saturaţie foarte mare într-un factor ("factorul primar") şi o saturaţie mare în alt factor ("factorul secundar"). Pentru a se determina ce măsoară factorii se grupează elementele care au acelaşi factor primar şi se analizează conţinutul acestora. Deosebirile dintre concluziile formulate de diverşi autori sunt cauzate de faptul că factorii secundari la elementele grupate pentru un acelaşi factor primar diferă de la un studiu la altul. De exemplu, o scală alcătuită din itemi care se referă la conştiinciozitate, dar care au saturaţii mari pozitive în factorul extraversiune, va conduce spre o altă denumire a factorului "conştiinciozitate" decât o scală compusă din itemi care vizează conştiinciozitatea dar au saturaţii mari negative în factorul extraversiune. În primul caz denumirea dată factorului "conştiinciozitate" va indica o formă extravertită a conştiinciozităţii, iar în al doilea caz, o formă introvertită.

Într-un model al persoanei, prezentat în 1994, de R.R. McCrae şi P.T. Costa, cei cinci factori din modelul Big Five sunt văzuţi ca dispoziţii psihice fundamentale, care au baze biologice, nu sunt observabile direct şi oferă explicaţii pentru o categorie de fapte psihice numite "adaptări psihice". Acestea din urmă sunt caracteristici dobândite şi includ obiceiuri, atitudini, deprinderi, valori şi motive. La rândul lor, adaptările caracteristice furnizează explicaţii pentru comportamentele observabile (gânduri, sentimente, acţiuni, adică ceea ce evaluează de obicei persoanele din jur) (Minulescu, 1996).

Unele cercetări au arătat că fiecare dintre cei cinci factori este legat de bunăstarea psihică şi de adaptarea din adolescenţă (Ozer, Reise, 1994).

Un model de evaluare a personalităţii acceptat de psihologi în ultimii ani constă dintr-o structură ierarhică: în vârf se găsesc cei cinci factori din modelul Big Five (numiţi, uneori, **superfactori**), iar sub aceştia sunt plasate diverse **faţete**, care sunt măsurate prin instrumente ce se referă la aspecte înguste ale personalităţii (Goldberg, 1996).

În ceea ce priveşte setul optim de faţete, psihologii nu au ajuns încă la un acord. De exemplu, în modelul AB5C propus de Hofstee, de Raad şi Goldberg există 45 de dimensiuni bipolare, în timp ce în modelul Five Factors al lui Costa şi Mc Crae, operaţionalizat în NEO-PI-R, sunt prezente 30 de dimensiuni bipolare.

În orice caz, sunt considerate mai utile chestionarele care cuprind scale corespunzătoare faţetelor decât cele care au doar scale de evaluare a superfactorilor (Goldberg, 1996). Aceasta, deoarece din primele se poate ajunge cu uşurinţă la superfactori, în timp ce drumul invers este imposibil.

Au fost construite mai multe liste de termeni şi chestionare care măsoară cei cinci superfactori ori faţete ale acestora.

S-au reţinut în liste: verbe (care indică "ce face o persoană"), adverbe (care arată "cum face" ceva persoana), adjective (care exprimă "de ce" se comportă persoana într-un anumit fel) şi substantive (care denumesc trăsăturile) (Zlate, 1994). Cea mai mare extindere o au listele alcătuite din adjective şi cele formate din substantive. Dar, se apreciază că asemenea descriptori compuşi dintr-un singur cuvânt nu sunt ideali pentru a fi utilizaţi ca itemi în chestionare de personalitate (Goldberg, 1996). Mai întâi, din cauza faptului că nici un limbaj nu cuprinde suficiente cuvinte pentru a surprinde toate nuanţele complexe în care este înţeleasă fiecare trăsătură. Apoi, pentru că adjectivele şi substantivele care se referă la trăsături au un nivel foarte înalt de abstractizare: două persoane înţeleg acelaşi termen în moduri diferite. De aceea, mulţi autori de chestionare preferă itemi care conţin specificaţii comportamentale sau contextuale. Pe de altă parte, nu totdeauna este posibil să se găsească o traducere potrivită în altă limbă a fiecărui termen printr-un singur cuvânt.

În general, itemii introduşi într-un chestionar de personalitate pot fi clasificaţi, din punct de vedere al aspectului vizat, astfel (Minulescu, 1996):

- după specificitatea comportamentului:
    - abstracţi (nespecifici);
    - concreţi (specifici);
- după specificitatea situaţională:
    - condiţionali (se referă la comportamente specifice pentru o anumită situaţie);
    - necondiţionali (nu sunt sensibili la specificul situaţional).

Pornind de la această clasificare, B. De Raad şi colaboratorii săi au găsit patru tipuri de itemi potriviţi instrumentelor de tip Big Five (Minulescu, 1996):

a. adjective singulare (de exemplu, extravert, stabil, conştient) care sunt simultan abstracte şi necondiţionale;

b. verbe singulare (de exemplu, se bâlbâie, roşeşte, insultă) care sunt concrete şi necondiţionale;

c. propoziţii adjectivale (de exemplu, este deschis când se află în prezenţa altora) care sunt abstracte şi condiţionale;

d. propoziţii cu centrul de greutate pe sensul verbului (de exemplu, se bâlbâie când este în prezenţa unei persoane de sex opus) care sunt concrete şi condiţionale.

În timp ce propoziţiile adjectivale apar frecvent în chestionarele de personalitate, propoziţiile centrate pe verb au fost rareori folosite.

Printre cele mai utilizate liste de adjective construite pentru măsurarea celor cinci superfactori se află:

- *50-BRS* (Bipolar Rating Scales), lista de adjective a lui L. R. Goldberg, care constă din 10 scale bipolare pentru fiecare superfactor;
- *SACBIF* (Short Adjective Checklist Big Five), o listă realizată de M. Perugini, L. Leone, L. Gallucci şi M. Lauriola, care cuprinde 50 de adjective bipolare pentru cele 5 scale referitoare la superfactori.

Dintre chestionarele de personalitate construite după modelul Big Five amintim:

- *NEO-PI-R* (NEO Personality Inventory Revised) al lui P.T. Costa şi R. R. McCrae, care măsoară câte şase faţete pentru fiecare superfactor. Există în circulaţie două variante ale chestionarului NEO-PI-R, ambele compuse din câte 240 de itemi: Forma S, pentru autoevaluare, şi Forma R, pentru eteroevaluare.
- *FFPI* (Five-Factor Personality Inventory) construit de A. A. J. Hendriks, W. K. B. Hofstee, B. d Raad şi A. Angleitner, care a fost realizat de la început în trei versiuni standardizate: olandeză, germană şi engleză. El constă din 100 de itemi, repartizaţi câte 20 pentru fiecare superfactor.

3.4.2. Modelul circumplex

Modelul circumplex a fost elaborat de J. S. Wiggins. El constă dintr-o reprezentare grafică a trăsăturilor interpersonale în interiorul unui cerc, în spaţiul bidimensional care are ca şi axe de coordonate Puterea (sau Dominanţa) şi Dragostea (sau Pasiunea). Cercul are centrul în originea axelor, raza egală cu unitatea şi este împărţit în 16 sectoare egale.

Wiggins a selectat, dintr-o listă cu 1710 adjective construită de Goldberg, 817 termeni care se referă la comportamente interpersonale. Apoi, a studiat corelaţiile dintre autoevaluările făcute de subiecţi la aceste adjective şi, pe baza lor, a grupat termenii în 16 clase, fiecare corespunzând unui sector de cerc. Rafinări succesive ale acestor grupe de adjective au dus la formarea a 16 scale, fiecare constând din 8 termeni.

Modelul circumplex a fost generalizat de Hofstee, De Raad şi Goldberg (1992), cercul fiind înlocuit cu o sferă în spaţiul cu cinci dimensiuni, în care axele de coordonate corespund superfactorilor din modelul Big Five. Noul model este denumit *AB5C* (*Abridged Big Five Dimensional Circumplex*).

În acest circumplex cu cinci dimensiuni fiecare trăsătură este caracterizată prin saturaţiile sale în doi din cei cinci superfactori. Modelul *AB5C* poate fi privit ca fiind format din 10 circumplexe bidimensionale, de tipul celui construit de Wiggins, câte unul pentru fiecare combinaţie posibilă de doi factori. În fiecare circumplex bidimensional pe axele *Ox* şi *Oy* sunt reprezentaţi superfactorii din perechea considerată. De data aceasta cercul este împărţit în 12 sectoare, fiecare având unghiul la centru egal cu 30°, poziţionate astfel încât axele de coordonate să fie bisectoare pentru patru dintre ele.

În acest circumplex sunt reprezentate numai trăsăturile care, după rotirea factorilor prin metoda varimax, au cele mai mari saturaţii în superfactorii cărora le corespund axele *Ox* şi *Oy*.

Algoritmul de atribuire a unei trăsături la un sector constă din următorii paşi (Hofstee, De Raad, Goldberg, 1992):

Pasul 1:

Se desenează punctul care are coordonatele rectangulare (x, y), unde x este saturaţia trăsăturii în superfactorul reprezentat pe axa *Ox*, iar y este saturaţia în superfactorul reprezentat pe axa *Oy*.

Pasul 2:

Se calculează lungimea proiecţiei punctului pe bisectoarea sectorului de cerc din care face parte, astfel:

- dacă $|x|>3.73|y|$ atunci punctul face parte dintr-un sector de cerc care are ca şi bisectoare axa *Ox* şi proiecţia sa pe această axă are lungimea $h=|x|$;
- dacă $|y|>3.73|x|$ atunci punctul face parte dintr-un sector de cerc care are ca şi bisectoare axa *Oy* şi proiecţia sa pe această axă are lungimea $h=|y|$;
- dacă nu sunt îndeplinite inegalităţile de mai sus, atunci punctul aparţine unui sector care nu conţine axele de coordonate şi proiecţia sa pe bisectoarea sectorului se calculează cu formula:

$$h = \max\{|x|,|y|\} \cos 30^\circ + \min\{|x|,|y|\} \cos 60^\circ \qquad (3.4.2.1)$$

Pasul 3

Trăsăturile pentru care proiecţia are lungimea mai mică decât un prag fixat (de exemplu, decât 0.20) sunt eliminate din circumplex.

După reprezentarea tuturor trăsăturilor, se analizează grupele din fiecare sector de cerc şi se caută denumirea cea mai potrivită pentru fiecare superfactor.

Deşi modelul *AB5C* nu diferă formal de modelul Big Five în ceea ce priveşte structura internă a personalităţii, el se deosebeşte de acesta prin faptul că introduce factori suplimentari printre factorii ortogonali: cei corespunzători bisectoarelor sectoarelor care nu conţin axele de coordonate (Johnson, 1993).

3.5. UTILIZAREA CALCULATOARELOR ÎN TESTAREA PSIHOLOGICĂ

Utilizarea calculatoarelor în domeniul testării psihologice are o vechime de peste 40 de ani. La început, ele erau folosite doar pentru păstrarea şi analizarea rezultatelor la teste în cazul loturilor mari de subiecţi. Dar, evoluţia tehnicii de calcul şi, mai ales, necesităţile izvorâte din practică au provocat o diversificare a funcţiilor programelor destinate psihologilor. De exemplu, apariţia primului program pentru interpretarea rezultatelor unui test a fost determinată de acumularea unui număr foarte mare de lucrări referitoare la MMPI. Devenise imposibil, chiar şi pentru psihologii cei mai competenţi, să memoreze toate informaţiile conţinute în acestea şi, cu atât mai mult, să le identifice rapid pe cele mai potrivite pentru interpretarea unui protocol particular.

Programele folosite azi în domeniul testării psihologice îndeplinesc una sau mai multe din următoarele funcţiuni:

a. *păstrarea rezultatelor testării psihologice*, în fişiere sau în baze de date. Prin aceasta se economiseşte hârtie şi se reduce considerabil timpul necesar pentru regăsirea informaţiilor despre o anumită persoană. Datele din fişiere pot servi la analizarea testelor (de exemplu, la verificarea validităţii acestora) sau la stabilirea unor etaloane.

b. *calculul scorurilor la teste*. Se introduc în calculator răspunsurile date de un subiect la itemii testului şi, pe baza grilei, programul determină cotele brute ale scalelor. Asemenea programe sunt utile mai ales în cazul testelor compuse dintr-un număr mare de itemi, repartizaţi pe mai multe scale, cum sunt MMPI şi CPI. Există şi teste care nu pot fi cotate decât cu ajutorul calculatorului (de exemplu, testul de capacitate de organizare CO92).

c. *administrarea testelor*, urmată, evident, de calculul scorurilor. Calculatorul poate administra teste exact în aceeaşi formă (aceiaşi itemi, în aceeaşi ordine) ca în varianta de administrare creion-hârtie sau poate "adapta" testele pentru fiecare subiect, extrăgând dintr-o colecţie de itemi pe cei care permit obţinerea unei informaţii cât mai precise despre persoana în cauză. Acest procedeu de lucru, denumit *testare adaptativă cu ajutorul calculatorului*

(Computerized Adaptive Testing, pe scurt, CAT) prezintă următoarele avantaje, comparativ cu testarea obişnuită:

- determinarea nivelului persoanelor în privinţa caracteristicii psihice măsurate este mai precisă;
- examinarea durează mai puţin, testele adaptative fiind compuse dintr-un număr mai mic de itemi, şi, prin urmare subiectul este solicitat mai puţin, iar costul testării este mai redus;
- securitatea testului este mai uşor de păstrat, întrucât fiecare persoană examinată cunoaşte (şi, deci, poate memora) doar un număr redus de itemi din totalul itemilor posibili.

În urma mai multor cercetări s-a constatat că forma de administrare a testului are influenţă asupra răspunsurilor date de subiecţi (Murphy, Davidshofer, 1991). Din acest motiv, se impune să se determine şi să se utilizeze etaloane separate pentru fiecare mod de administrare a testului.

Cu ajutorul calculatorului pot fi administrate teste prin care se măsoară variabile psihice sau fiziologice ce nu ar putea fi investigate în cazul administrării creion-hârtie (de exemplu, timpul de reacţie, timpul de decizie, timpul de răspuns la fiecare item al unui chestionar etc.).

d. *identificarea protocoalelor invalide*, în care răspunsurile subiectului nu sunt conforme cu realitatea (au fost date la întâmplare sau sunt influenţate de dorinţa persoanei examinate de a apărea într-un anumit fel). Un procedeu utilizat în acest scop aplică teste statistice asupra succesiunilor de răspunsuri date de subiect (Huba, 1987).

e. *verificarea unor ipoteze referitoare la persoana examinată*, pe baza comparării, prin teste statistice, a rezultatelor obţinute de aceasta la examenul psihologic cu cele ale unui eşantion extras din populaţia căreia îi aparţine subiectul.

f. *intervievarea subiectului.* Programele conţin, alături de întrebările posibile, şi un algoritm de construire a interviului, în timpul examenului psihologic, în funcţie de răspunsurile date de subiect. Interviurile realizate de calculator sunt mai flexibile decât chestionarele cu structură fixă. În plus, se pare că ele aduc informaţii mai precise decât interviurile conduse de psiholog, întrucât s-a constat că indivizii sunt, în general, mai dispuşi să comunice informaţii despre sine unei maşini, în particular unui calculator, decât unei alte persoane (Aiken, 1994). Dar, interviurile administrate de calculator au o utilitate redusă în cazul copiilor, al adulţilor cu un nivel intelectual scăzut şi al celor cu simptome psihiatrice.

g. *interpretarea rezultatelor la un test psihologic*: Interpretarea testului de către calculator (Computer-based Test Interpretation, pe scurt, CBTI) are la bază transpunerea într-un program a unui set de reguli prespecificate, referitoare

la un răspuns sau la un pattern de răspunsuri (un scor de test sau un profil psihologic), care permit analiza, interpretarea şi evaluarea unor calităţi ale persoanelor (Harris, 1987). Cea mai simplă variantă de interpretare constă în transformarea cotelor brute în scoruri standardizate (cote *z* sau cote *T*) ori în ranguri percentile (decile, cuartile etc.) sau în calculul intervalelor de încredere pentru fiecare scor. Pe o treaptă superioară se află interpretările compuse din înşiruiri de adjective, iar la nivelul cel mai înalt, interpretările alcătuite din propoziţii. Atunci când regulile aflate la baza algoritmului utilizat de program au fost emise de experţi sau au fost deduse din analiza statistică a unui număr foarte mare de rapoarte elaborate de psihologi competenţi, interpretările realizate de calculator sunt foarte asemănătoare celor produse de specialişti şi, în orice caz, sunt superioare celor formulate de un psiholog începător (Huba, 1987). Programele bune iau în considerare, alături de scorurile realizate de subiect, şi diverse date biografice. Însă, ele prezintă neajunsul că nu pot utiliza informaţii accesibile doar prin observaţie, cum sunt mimica sau gestica.

h. *redactarea raportului psihologic.* Sunt prezentate explicaţii ale scorurilor obţinute la teste de persoana examinată sau implicaţii posibile ale acestor rezultate. Uneori sunt formulate predicţii referitoare la subiect, fie pe baza unor metode statistice (de exemplu, folosind regresia liniară), fie pe baza identificării unor legături între scorurile la test şi unele caracteristici non-test, cum sunt datele biografice. Alteori, în urma comparării profilului psihologic al subiectului cu rezultatele obţinute la aceleaşi teste de diverse grupuri de persoane (de exemplu, grupuri care diferă între ele prin profesie, prin performanţa în muncă sau prin diagnosticul psihiatric) se determină populaţia din care face parte persoana examinată.

i. *alegerea tratamentului (a terapiei, a programului de instruire etc.) cel mai potrivit pentru fiecare subiect.* Acesta este cel mai înalt nivel de implicare a calculatorului în activitatea psihologului. Dar, programele de acest tip sunt rare deocamdată.

j. *construirea unui test psihologic.* Sunt analizate, prin metode statistice, răspunsurile date la itemii testului de un lot numeros de persoane şi sunt identificaţi acei itemi care, puşi laolaltă, alcătuiesc un test cu calităţi cât mai bune.

Fiind uşor de utilizat, programele de testare psihologică au cunoscut o mare răspândire. Mulţi consideră, în mod eronat, că asemenea programe pot înlocui psihologul sau că, la fel ca şi programele de jocuri, pot fi folosite oricând şi de oricine.

În *Standards for Educational and Psychological Testing* (1985) şi în *Guidelines for Computer-Based Tests and Interpretations* (1986) sunt precizate o serie de condiţii în care pot fi utilizate programele de testare psihologică.

Unele dintre acestea se referă la aspectele tehnice ale testării, iar altele, la obligaţiile care îi revin persoanei care le foloseşte.

De exemplu, în legătură cu administrarea testului, *Guidelines for Computer-Based Tests and Interpretations* formulează următoarele cerinţe (Aiken, 1994):

a. calculatorul utilizat la testare să fie plasat într-un loc liniştit, confortabil şi lipsit de factori care să distragă atenţia;

b. itemii prezentaţi pe ecran să fie lizibili, iar strălucirea ecranului să fie potrivită;

c. echipamentul să fie verificat periodic şi să funcţioneze bine;

d. persoana examinată să fie instruită în privinţa modului de utilizare a calculatorului;

e. să i se ofere asistenţă persoanei examinate ori de câte ori este necesar şi să se urmărească comportarea acesteia la test;

f. persoanelor care ar putea fi dezavantajate prin testarea cu ajutorul calculatorului să li se asigure o perioadă de acomodare.

Standards for Educational and Psychological Tests din 1985 menţionează că recomandările referitoare la teste sunt valabile şi în cazul testării cu ajutorul calculatorului. Există, în plus, unele recomandări specifice, majoritatea vizând utilizarea interpretărilor furnizate de calculator.

Mulţi specialişti în domeniul testării psihologice, la fel ca şi diverse standarde elaborate de APA, atrag atenţia că programele de interpretare a testelor au doar rolul de a-l ajuta pe psiholog şi nu de a-l înlocui. Persoanele care folosesc un asemenea program trebuie să cunoască bine testul la care se referă acesta, să fie capabile ca, luând în considerare contextul în care s-a efectuat examenul psihologic şi diverse caracteristici ale subiectului, să recunoască dacă interpretarea propusă de calculator este corectă sau nu. De altfel, încă din 1966, în *Automated test scoring and interpretation practices*, APA atrăgea atenţia că realizarea interpretării unui test de către calculator trebuie privită ca o "consultaţie de la profesionist la profesionist", similară celor oferite de psihologi studenţilor în psihologie sau altor persoane pe care le supervizează.

CAPITOLUL 4
CLASIFICĂRI ALE TESTELOR PSIHOLOGICE

4.1. CLASIFICĂRI ÎN FUNCŢIE DE CONŢINUTUL TESTULUI

A. După *modul de executare a sarcinii de către subiect*, testele se clasifică în:
- *teste orale*;
- *teste scrise* ("creion-hârtie");
- *teste de performanţă*:
  - teste cu manipulare de aparate şi piese;
  - teste administrate de calculator.

În cazul testelor "creion-hârtie", itemii sunt scrişi pe hârtie, iar subiecţii răspund în scris, fie pe foaia care conţine testul, fie pe o foaie de răspuns. Există tendinţa de înlocuire a acestor teste cu testele administrate de calculator. Itemii nu mai apar pe hârtie, ci sunt afişaţi pe ecranul monitorului, iar subiectul nu mai răspunde pe hârtie, ci apasă o tastă sau manevrează mouse-ul ori creionul optic.

Testele de performanţă impun subiecţilor să opereze cu obiecte, desene, aparate sau să efectueze activităţi complexe, asemănătoare situaţiilor din viaţa reală. În această categorie sunt incluse şi testele în care stimulii sunt constituiţi din desene animate, prezentate pe un ecran, la televizor sau pe monitorul unui calculator.

Spre deosebire de testele creion-hârtie, care, de obicei, se administrează colectiv, testele de performanţă sunt utilizate individual sau în grupuri restrânse, atât datorită dificultăţii de a avea multe exemplare din materialele sau aparatele cu care persoanele examinate trebuie să lucreze, cât şi pentru a evita ca subiecţii să vadă ce realizează persoanele din jur.

În ultimii ani, a cunoscut o mare extindere examinarea psihologică cu ajutorul calculatorului.

B. După *gradul de utilizare a limbajului în rezolvarea sarcinii impuse de test* se poate vorbi despre:
- *teste verbale* (de exemplu, teste de raţionament verbal, teste de vocabular, teste care cer completarea unor propoziţii);
- *teste non-verbale* (de exemplu, teste care cer urmărirea mintală a unui labirint, rotirea pe plan mintal a unei figuri, completarea unor figuri lacunare).

În cazul testelor non-verbale nu se utilizează limbajul nici oral, nici în scris, nici în conţinutul testului şi nici în instrucţiunile acestuia. Asemenea teste

sunt destinate persoanelor analfabete, cu deficienţe de auz sau care nu cunosc limba folosită la examinare.

Unii autori acceptă sub denumirea de teste non-verbale şi probe în care utilizarea limbajului este redusă. Pentru asemenea teste au fost stabilite următoarele trei niveluri (English, English, 1970):

- teste complet non-verbale (extrem de rare);
- teste non-verbale în instrucţiuni, în forma de prezentare a sarcinii şi în modul de răspuns din partea subiectului, dar care acceptă utilizarea limbajului pe parcursul efectuării sarcinii (de exemplu, subiectul poate comenta ceea ce face, vorbind singur);
- teste în care numai răspunsul subiectului este non-verbal.

Un test non-verbal poate fi test de performanţă sau test creion-hârtie. În ultimul caz testul constă din desene, diagrame sau simboluri non-lingvistice, iar subiectul răspunde prin efectuarea unor marcaje simple. Instrucţiunile sunt date prin gesturi sau pantomimă, iar pentru demonstraţii se folosesc planşe, tabele şi diagrame.

Uneori denumirea de "test non-verbal" este sinonimă cu aceea de "test de performanţă" şi desemnează un "test care cere răspunsuri motorii şi nu verbale" (English, English, 1970). Alteori clasificarea după gradul de utilizare a limbajului face distincţie între "testele verbale", "testele non-verbale" (care utilizează material figural) şi "testele de performanţă" (care utilizează diverse materiale concrete) (Roşca, 1972).

C. După *modul de procesare implicat* (ce i se cere persoanei să facă pentru a da un răspuns) testele se clasifică în:

- *teste de eficienţă* (teste intelective):
  - teste de aptitudini;
  - teste de dezvoltare mintală sau de inteligenţă;
  - teste de cunoştinţe;
  - probe de lucru;
  - teste situaţionale;
- *teste de personalitate* (teste nonintelective):
  - chestionare de personalitate;
  - tehnici (teste) proiective;
  - teste obiective de personalitate.

Unii autori includ testele de inteligenţă în categoria testelor de aptitudini (Reber, 1985).

În mod obişnuit, sub denumirea de teste de personalitate sunt cunoscute teste care se ocupă cu aspecte afective ori nonintelectuale sau cu comportamente. Unii psihologi cuprind în această categorie teste care se referă atât la

trăsături nonintelectuale, cât şi la trăsături intelectuale (Anastasi, 1976). În terminologia testării psihologice, totuşi, prin teste de personalitate se înţeleg cel mai frecvent măsurători ale unor caracteristici precum adaptarea emoţională, relaţiile interpersonale, motivaţia, trebuinţele fundamentale, interesele, atitudinile (Anastasi, 1976; Tyler, 1963). Adesea, itemii lor se bazează pe expunerea unor convingeri sau atitudini ("Eu cred că ..."; "Un lider ar trebui să ..."; Mă simt ...").

În clasificarea prezentată mai sus, împărţirea în cele două grupe mari (teste de eficienţă şi teste de personalitate) corespunde şi nivelului de performanţă evaluat de teste. Astfel, testele de eficienţă măsoară performanţa maximă a unei persoane, în timp ce testele de personalitate se referă la performanţa tipică (Corsini, 1994).

Testele care evaluează performanţa maximă sunt compuse din itemi la care se răspunde corect sau greşit. Ele furnizează de obicei un scor numeric, care permite ierarhizarea subiecţilor.

În probele de personalitate nu există răspunsuri "bune" sau "rele". Ele evaluează în general diferenţe de stil între persoane. Pe baza rezultatelor lor nu se poate afirma că un subiect este, în privinţa tuturor variabilelor măsurate de test, mai bun sau mai slab decât altul.

● **Testele de aptitudini**

Acestea sunt probe care oferă informaţii despre gradul de dezvoltare a variabilelor intelectuale, psihomotorii şi psihofiziologice (creativitate, coordonare manuală, percepţie cromatică). Ele permit construirea unui profil individual diferenţiat şi stabilirea unei ierarhii între persoane.

O clasificare a acestor probe cuprinde (Reber, 1985):

- *teste de aptitudini speciale*: sunt cele care măsoară potenţialul pentru o singură aptitudine (de exemplu, mecanică, muzicală etc.);
- *teste de aptitudini generale*: ele determină potenţialul în domenii nespecifice (testele de inteligenţă fac parte din această categorie);
- *teste de aptitudini multiple*: sunt baterii de teste care evaluează mai mulţi factori.

Această clasificare nu este precisă. De exemplu, testul de dexteritate a degetelor este "special" pentru o aptitudine specifică, dar măsoară un aspect al unei aptitudini generale senzorio-motorii. Unele teste de inteligenţă, care sunt teste de aptitudini generale, au fost prelucrate prin analiză factorială şi s-au pus în evidenţă componente verbale, numerice, analitice, spaţiale etc., ce ţin de aptitudini particulare, speciale.

● **Testele de inteligenţă**

Sunt denumite şi "teste de măsurare a dezvoltării intelectuale generale" (Roşca, 1972). Ele exprimă rezultatul prin intermediul unui coeficient de vârstă mintală sau de vârstă de dezvoltare.

Există teste de inteligenţă verbale, teste cu conţinut figurativ şi teste în care rezolvarea constă în manipularea unui material concret. Testele de inteligenţă pur verbale sunt rare; majoritatea testelor care sunt predominant verbale conţin şi probe non-verbale.

Exceptând testele de cunoştinţe utilizate în şcoli, testele de inteligenţă sunt testele folosite cel mai frecvent.

● **Testele de cunoştinţe**

Acestea constituie un mijloc de apreciere a fondului de informaţii al persoanei.

Adesea se face distincţie între testele de cunoştinţe şi testele de aptitudini. În timp ce testele de cunoştinţe îi cer persoanei examinate să furnizeze o informaţie (de exemplu, să spună cine a înfiinţat primul laborator de psihologie experimentală), testele de aptitudini îi pretind să efectueze o sarcină (de exemplu, pentru testarea raţionamentului numeric, să spună câte mere va avea fiecare copil dacă se împart în mod egal 81 de mere la 27 de copii) (Hammond, 1995).

Testele de cunoştinţe evaluează cantitatea de informaţii pe care a dobândit-o subiectul din experienţa anterioară, în timp ce testele de aptitudini măsoară cât de mult este capabilă o persoană să rezolve o problemă concretă cu care nu a mai avut de-a face până atunci (Hammond, 1995). Dar această distincţie bazată pe experienţa anterioară nu este valabilă totdeauna, deoarece testele de aptitudini generale implică utilizarea unor principii învăţate anterior.

Deosebirea principală dintre testele de aptitudini şi cele de cunoştinţe este dată de direcţia în care este orientat interesul examinatorului. El aplică un test de aptitudini dacă doreşte să afle ce *poate învăţa* o persoană şi utilizează un test de cunoştinţe dacă vrea să ştie *ce a învăţat* individul până în momentul examinării (Thorndike, Hagen, 1961).

Dar, ceea ce o persoană a înfăptuit în trecut este, de obicei, un indicator bun pentru ceea ce va putea ea realiza în viitor. Din acest motiv, unii psihologi preferă ca în locul denumirilor de "test de aptitudini" şi "test de cunoştinţe" să utilizeze termenul de "*test de abilităţi*" (Aiken, 1994).

Deşi din definiţia testului se deduce că acesta este o probă standardizată, în cazul testelor de cunoştinţe se obişnuieşte să se vorbească despre (Gronlund, 1967):

- *teste standardizate* (elaborate de specialişti, administrate, cotate şi interpretate în condiţii standardizate, destinate să fie utilizate pe scară largă);
- *teste informative* (elaborate de profesorul care predă la o clasă şi interpretate prin raportare la scorurile clasei respective).

Deosebirea principală dintre testele standardizate şi cele informative este dată de scopul pentru care sunt folosite. Testele elaborate de către profesor se folosesc în primul rând pentru măsurarea îndeplinirii obiectivelor imediate din clasă şi, deci, se administrează grupului pentru care au fost întocmite. Testele standardizate se utilizează pentru a compara activitatea unui grup cu cea a altuia, pentru a compara rezultatele unei clase sau ale unui individ cu datele externe oferite de normele testului. Aceste deosebiri de scop determină diferenţe de structură, de administrare şi de interpretare între cele două tipuri de teste.

Testele standardizate se deosebesc de cele informative din trei puncte de vedere (Lindeman, 1978):

a. Testele standardizate sunt administrate totdeauna în aceleaşi condiţii. Fiecare test standardizat este însoţit de un manual, care include instrucţiuni pentru administrare şi cotare. Timpul alocat diverselor subsecţiuni ale testului este menţionat în aceste instrucţiuni şi trebuie respectat cu stricteţe, pentru a putea interpreta rezultatele prin comparare cu cele obţinute de alt grup sau de altă persoană.

b. Testele standardizate cuprind totdeauna norme, determinate pe baza unei populaţii numeroase.

c. Testele standardizate acoperă o mai mare varietate de conţinut decât cele informative. Ele pot fi inadecvate pentru măsurarea unor obiective restrânse din anumite şcoli.

O altă clasificare a testelor de cunoştinţe are la bază funcţia acestora. Se deosebesc, astfel (Gronlund, 1967; Guilbert, 1987):

- *teste cu criterii absolute* (mastery tests), care verifică dacă subiecţii posedă cunoştinţele absolut necesare pentru un anumit scop (promovarea clasei, ocuparea unui post etc.); ele evaluează performanţa unui individ în raport cu un nivel de performanţă specificat dinainte;
- *teste cu criterii relative* (survey tests), care servesc la ierarhizarea subiecţilor în funcţie de cunoştinţele acestora;
- *teste diagnostice* (diagnostic tests), care au rolul de a identifica deficienţele subiecţilor, cunoştinţele absente sau incomplete.

Deci, testele cu criterii absolute se referă la performanţa minimă (subiecţii trebuie să realizeze cel puţin atât cât prevede criteriul absolut), testele cu criterii relative se referă la performanţa maximă (subiecţii sunt ierarhizaţi în

funcţie de scorul obţinut la test), iar testele diagnostice urmăresc depistarea domeniilor în care performanţa este foarte scăzută.

• Probele de lucru

Probele de lucru (work sample tests) sunt teste de performanţă bazate pe elemente care au legătură cu munca pe care o vor efectua subiecţii. Ele măsoară deprinderile implicate în toate aspectele importante ale muncii respective. Testele de acest fel se utilizează mai ales în cazul profesiilor care necesită manipulări de obiecte, fie pentru a măsura ceea ce a învăţat subiectul, fie pentru a prezice ceea ce va putea învăţa (Downs, 1989).

Unele probe de lucru (trainability worksample tests) se administrează înaintea unui program de instruire - a cărui lungime poate varia între o săptămână şi şase luni -, când persoanele nu sunt încă familiarizate cu activităţile pe care ar urma să le efectueze după ocuparea postului, pentru a prezice dacă subiecţii îşi vor însuşi deprinderile pe care respectivul program urmăreşte să le formeze. În majoritatea cazurilor, administrarea acestor teste se face după efectuarea selecţiei profesionale.

Alte probe de lucru (job sample test) se utilizează după ocuparea postului de către subiecţi, pentru evaluarea deprinderilor însuşite de aceştia. Ele conţin eşantioane din munca efectuată la postul respectiv. De exemplu, pentru un post de secretară, o asemenea probă poate consta din editarea unui text la calculator.

• Testele situaţionale

Testele situaţionale sunt echivalentul probelor de lucru în cazul profesiilor administrative. Ele permit evaluarea unor caracteristici care nu pot fi măsurate de testele cognitive. Dar, în general, sunt costisitoare şi consumă mult timp. Aceste teste sunt utilizate mai ales pentru selectarea managerilor.

Principala deosebire dintre o probă de lucru şi un test situaţional constă în faptul că proba de lucru se referă doar la o parte a sarcinii care îi revine postului, în timp ce testul situaţional oglindeşte întreaga sarcină (P. M. Muchinsky, după Gregory, 1992). Probele de lucru pun accentul pe evaluarea de aptitudini şi de cunoştinţe, iar testele situaţionale se ocupă mai mult cu variabile emoţionale, sociale, atitudinale şi de personalitate.

Testele situaţionale îl pun pe subiect în condiţii asemănătoare celor în care ar lucra dacă ar fi acceptat la selecţie şi determină cât de potrivit este răspunsul său. Scopul acestor teste este deghizat, persoana care îl rezolvă nu îşi poate da seama care aspecte ale performanţei sale sunt observate şi cotate.

Un test situaţional utilizat în selecţia profesională le poate oferi persoanelor examinate câteva probleme "reale" care cer îndeplinirea unei sarcini

în condiţii de limitare a resurselor necesare (Gilmer, 1971). De exemplu, pentru ocuparea unui post de conducere al unui colectiv, candidatului i se poate pune la dispoziţie o echipă care să îl ajute în îndeplinirea sarcinii. Dar, membrii echipei sunt instruiţi în prealabil să împiedice desfăşurarea activităţii, cât mai mult posibil. Se urmăreşte şi se cotează comportamentul celui examinat: cum ştie să îi conducă pe cei "incompetenţi", cum reacţionează la frustrare şi în ce măsură este capabil să ducă la bun sfârşit sarcina impusă de test.

Un alt test situaţional este tehnica "in-basket", care simulează condiţiile de lucru ale unui administrator. Testul clasic in-basket este *The Bureau of Business In-Basket Test* al lui N. Frederikson (Gregory, 1992), care constă din scrisori, însemnări în agenda de lucru, notiţe despre apeluri telefonice şi alte documente care au fost adunate în coşul de pe masa de lucru a unui director la o firmă care se ocupă de afaceri. Subiectului i se cere să îşi imagineze că trebuie să ţină locul directorului şi să îndeplinească sarcinile acestuia. El va trebui să dea telefoane, să alcătuiască scrisori, să răspundă la interviuri. Se apreciază atât conţinutul, cât şi stilul răspunsurilor sale, urmărindu-se şi modul în care subiectul ştie să colaboreze cu personalul firmei.

- **Chestionarele de personalitate**

Chestionarele de personalitate sunt compuse din una sau mai multe scale. O **scală de personalitate** este alcătuită dintr-o mulţime finită de declaraţii descriptive, numite **itemi**, care sunt relevante pentru o anumită trăsătură comună, adică pentru o anumită dimensiune a personalităţii, prezentă în grade diferite la indivizi diferiţi (de exemplu, anxietatea, agresivitatea) (Edwards, 1970).

După conţinutul aspectelor cercetate, chestionarele de personalitate pot fi clasificate astfel (Roşca, 1972):

- *chestionare de adaptare*, numite uneori şi *metode de autoevaluare* (self rating sau self report methods), deoarece subiectul dă informaţii asupra propriei conduite, a propriilor trăiri, care ar fi indicatoare pentru adaptarea sau pentru sănătatea sa psihică; se mai numesc şi *inventare de personalitate*;
- *chestionare de interese*;
- *chestionare de atitudini.*

După întinderea ariei aspectelor vizate, chestionarele de personalitate sunt (Roşca, 1972):

- *chestionare unifazice*, care explorează componentele unei singure variabile (de exemplu, chestionarul de ascendenţă-supunere al lui Allport);

• *chestionare multifazice*, care au în vedere mai multe variabile (de exemplu, Chestionarul de personalitate EPQ al lui Eysenck, care conţine scale pentru evaluarea introversiunii-extraversiunii, neuroticismului, psihoticismului şi a dezirabilităţii sociale).

Chestionarele de personalitate constau, în general, din afirmaţii sau întrebări, iar persoanei examinate i se cere să îşi manifeste acordul sau dezacordul faţă de ele. Acestea sunt formulate astfel încât, prin răspunsurile date, subiectul dă informaţii asupra stării sale de sănătate fizică sau psihică şi a comportamentelor sale specifice în situaţii diverse.

În mod obişnuit, itemii unei scale care evaluează o trăsătură psihică conţin afirmaţii despre (Angleitner, John, Löhr, 1986):

a. reacţii observabile (ex. "Merg des la petreceri."), neobservabile (ex. "Mă gândesc mult la mine însumi.") sau simptome (ex. "Adesea simt un nod în gât.");

b. atribute, exprimate obişnuit prin adjective (ex. "Sunt un bun organizator.");

c. dorinţe şi interese (ex. "Uneori îmi place să particip la competiţii sportive.");

d. fapte biografice (ex. "În copilărie chiuleam adesea de la şcoală.");

e. atitudini şi credinţe (ex. "Cred că legea trebuie respectată cu stricteţe.");

f. reacţiile altora faţă de subiect (ex. "La petreceri sunt totdeauna în centrul atenţiei.");

g. comportamente neobişnuite, chiar anormale (ex. "Cineva încearcă să mă otrăvească.").

Elementele componente ale unui chestionar (afirmaţii, întrebări) care evaluează o aceeaşi stare sau trăsătură psihică sunt grupate într-o scală al cărei nume conţine denumirea aspectului măsurat (scală de anxietate, scală de agresivitate etc.)

Chestionarele de personalitate au la bază presupunerea că cea mai bună cale de a afla ceva despre un individ este de a-l întreba direct (Hammond, 1995). Dar, nu totdeauna răspunsurile date de subiect îl caracterizează corect. Cauza o poate constitui atât testul (există deficienţe în modul de formulare a itemilor şi, din acest motiv, subiecţii răspund la întâmplare sau nu aleg răspunsul potrivit pentru ei) cât şi subiectul (doreşte să facă o impresie bună sau proastă şi răspunde în consecinţă sau are tendinţa de a da acelaşi răspuns -"da" sau "nu"- la toţi itemii).

Chestionarele de personalitate sunt foarte mult utilizate datorită faptului că sunt uşor de administrat şi de cotat.

Dar, nu orice construct psihologic poate fi măsurat prin chestionare de personalitate. Spre exemplu, popularitatea nu poate fi evaluată în acest mod, ci numai urmărind reacţiile şi aprecierile celorlalţi (Ozer, Reise, 1994).

• **Testele proiective**

Ele permit cunoaşterea subiectului prin ceea ce îi este specific şi reprezintă o abordare a personalităţii în ansamblul ei. Această cunoaştere este realizată indirect.

Testele proiective sunt utilizate frecvent pentru identificarea componentelor de personalitate legate de funcţionări psihice anormale, de exemplu pentru a examina aspecte ale persoanelor considerate a fi inconştiente (Hammond, 1995). În acest sens, ele sunt folosite în psihiatrie şi în psihologia clinică, pentru precizarea diagnosticului, pentru fixarea direcţiei de desfăşurare a psihoterapiei şi pentru urmărirea efectelor măsurilor terapeutice. Ele oferă de asemenea sugestii asupra momentelor istoriei personale care trebuie să fie luate în considerare în cadrul convorbirilor, al observaţiei etc.

Testele proiective constau din prezentarea unor stimuli nestructuraţi sau slab structuraţi, ambigui sau incompleţi, care permit astfel obţinerea unei mari varietăţi de răspunsuri. Sarcina trasată prin instrucţia prealabilă are un caracter foarte general, dând libertate imaginaţiei subiectului. De exemplu, subiectului i se arată o serie de stimuli ambigui sub forma unor pete de cerneală şi i se cere să spună ceea ce îi trece prin minte sau i se prezintă câte un desen, pe marginea căruia trebuie să compună o poveste.

Metodele proiective au la bază ideea că subiectul, încercând să structureze materialul prezentat, să îi găsească semnificaţia sau să îl completeze, "proiectează" asupra stimulului propriile sale interese, aspiraţii, atitudini, conflicte sau chiar moduri stabile de gândire (Roşca, 1972). Proiecţia efectuată de subiect este condiţionată de urmele lăsate în memoria acestuia de percepţiile anterioare (Petrovsky, Yaroshevsy, 1985).

Unii autori consideră că denumirea de tehnici proiective este inadecvată, deoarece în timpul executării unei astfel de probe nu intră în acţiune mecanismul de proiecţie în sensul utilizat de psihanalişti, mecanism care s-ar realiza inconştient, constituind o modalitate de descărcare a tensiunii. Ei propun pentru acest tip de tehnici denumirile *teste de apercepţie*, *teste de dinamism* sau *teste nestructurate*.

Unul dintre avantajele tehnicilor proiective constă în faptul că ele reduc rolul simulării, întrucât nu îi dau posibilitate subiectului să ghicească modul în care sunt interpretate răspunsurile.

Dar, psihotehnicienii au, în general, o atitudine critică faţă de probele proiective, considerând că ele nu îndeplinesc condiţiile unui instrument

standardizat de diagnostic. Administrarea, cotarea şi interpretarea acestor teste este laborioasă, cere un personal cu practică îndelungată, deci ele sunt costisitoare. În plus, interpretarea dată de examinator este adesea subiectivă şi arbitrară.

O altă deficienţă a testelor proiective este faptul că ele operează la nivelul nominal de măsurare, deci furnizează o descriere a persoanei examinate sub forma apartenenţei la o categorie. Pentru testele proiective folosite frecvent există şi proceduri de cuantificare a răspunsurilor individuale, dar aplicarea lor este dificilă, pentru că numărul variantelor de răspuns este extrem de mare.

- **Testele obiective de personalitate**

Aceste teste se aseamănă cu probele proiective, dar folosesc stimuli mai structuraţi.

Fiind teste obiective, ele urmăresc obţinerea unei concordanţe mari între cotările făcute de mai mulţi evaluatori. Stimulul şi reacţia provocată de el trebuie să fie de aşa natură încât subiectul să nu fie tentat să falsifice răspunsul sau să îl distorsioneze, din dorinţa de a se prezenta într-o lumină favorabilă sau nefavorabilă. Pentru a realiza o asemenea cerinţă testele sunt "deghizate"; cu alte cuvinte, subiectul nu-şi poate da seama ce trăsătură de personalitate se deduce din răspunsurile sale. Din acest motiv, testele obiective de personalitate sunt uneori denumite *teste indirecte* (Roşca, 1972).

Ipoteza care stă la baza construcţiei testelor obiective de personalitate este aceea că "stilul personalităţii", "stilul comportamentului" se manifestă în activităţi diverse, care nu seamănă unele cu altele. Sunt folosite în teste asemenea activităţi simple, manipulabile în mod experimental, care pot avea o valoare predictivă prin modul în care sunt îndeplinite. În general, subiectului i se pretinde să execute o sarcină, nu să se autoevalueze, şi, din felul în care acesta acţionează, se deduc diverse aspecte ale personalităţii.

De exemplu (Roşca, 1972), proba "Adjective check list" a lui H. G. Gough îi cere subiectului ca, dintr-o listă cu 300 de adjective aranjate în ordine alfabetică, să le marcheze pe cele care îl caracterizează. Psihologul are în vedere faptul că răspunsurile persoanei sunt conforme imaginii pe care aceasta o are despre sine şi nu realităţii, deci nu le interpretează direct. Dar, anumite aspecte ale personalităţii subiectului rezultă din modul de execuţie a sarcinii. S-a constatat, de pildă, că extravertiţii realizează aprecieri superioare celor date de introvertiţi, la trăsături care în realitate nu îi diferenţiază. Asemenea observaţii stau la baza interpretării testelor obiective de personalitate.

D. După ***modul în care variază proba de la o administrare la alta*** testele folosite într-o examinare psihologică se împart în:

- *teste cu conţinut fix*;
- *teste cu conţinut variabil*.

În primul caz, toate persoanele examinate cu ajutorul testului parcurg aceiaşi itemi. În general şi ordinea în care sunt prezentaţi itemii testului este aceeaşi. Atunci când testele sunt administrate colectiv, de calculator, pentru a se evita ca subiecţii să copieze răspunsurile unul de la altul, se poate recurge la schimbarea ordinii itemilor în cadrul testului, aleator sau după un algoritm stabilit a priori.

Majoritatea testelor utilizate în examinările psihologice au conţinut fix.

Din categoria testelor cu conţinut variabil fac parte *testele idiografice* şi testele *adaptative*.

În primul caz, testul este construit sau ales special pentru subiect.

Testele adaptative sunt generate în momentul examinării psihologice, în funcţie de comportarea subiectului.

După *Standards for Educational and Psychological Tests* din 1985, *testarea adaptativă* este "o formă secvenţială de testare în care itemii succesivi ai testului sunt aleşi pe baza răspunsurilor la itemii precedenţi". La fiecare pas se urmăreşte să se aleagă itemii cei mai potriviţi pentru a forma un test valid şi fidel, care să măsoare cu precizie în zona în care se află nivelul subiectului.

Formarea testului se realizează prin selectarea itemilor dintr-o colecţie existentă, compusă din itemi ale căror caracteristici sunt cunoscute dintr-o testare anterioară. Pe parcursul examinării psihologice testul se adaptează la nivelul de aptitudini sau de cunoştinţe al subiectului, realizând o măsurare cât mai precisă. Evident, asemenea teste nu se pot administra prin metoda creion-hârtie, ci doar cu ajutorul unor dispozitive speciale sau al calculatorului electronic.

Testele adaptative computerizate sunt construite, în general, aplicând teoria răspunsului la itemi. Selecţia itemului administrat la fiecare pas se face utilizând funcţiile de informaţie ale itemilor, astfel încât testul să furnizeze cât mai multă informaţie despre cel examinat (Stocking, 1987). Asemenea teste sunt administrate de calculator, conform următorului algoritm ((Weiss, Vale, 1987; Schoonman, 1989; Waller, Reise, 1989):

(1) Se alege un nivel de pornire pentru caracteristica măsurată de test. Acesta poate fi egal cu nivelul mediu în populaţia din care face parte subiectul.
(2) La fiecare pas al algoritmului:
- Se alege un item din baza de date.
- Se administrează itemul şi se preia răspunsul subiectului.
- Se estimează nivelul variabilei măsurate de test, în funcţie de răspunsurile date de subiect până în momentul respectiv.

- Se verifică dacă sunt îndeplinite condiţiile de oprire a testării; în caz afirmativ se trece la pasul 3, iar în caz negativ se reia de la pasul 2. Testarea se termină de obicei când a fost administrat numărul de itemi fixat, când s-a epuizat timpul de testare, când intervalul de încredere calculat pentru variabila măsurată de test are lungimea mai mică decât o valoare fixată sau când acest interval nu mai include pragul de admisibilitate fixat.

(3) Se comunică subiectului nivelul determinat de test.

Algoritmul prezentat se referă la cazul când toţi itemii testului măsoară o aceeaşi caracteristică. Dar, în teoria răspunsului la itemi au fost dezvoltaţi algoritmi care permit evaluarea unei persoane în privinţa mai multor variabile psihice.

La început, prin teste adaptative se evaluau doar aptitudini şi cunoştinţe. Acum se construiesc asemenea teste şi pentru măsurarea unor variabile non-cognitive, cum sunt atitudinile.

4.2. CLASIFICĂRI ÎN FUNCŢIE DE MODUL DE ADMINISTRARE A TESTULUI

A. După ***numărul persoanelor care pot fi examinate simultan***, testele pot fi:

- *individuale*;
- *colective* sau *de grup*.

Examenul psihologic individual este considerat neeconomic, subiecţii fiind testaţi unul câte unul, ceea ce necesită mult timp. Testarea colectivă presupune examinarea simultană a mai multor subiecţi.

Această clasificare nu este precisă, întrucât unele teste de grup pot fi aplicate şi în mod individual (de exemplu, testul Raven). Situaţia inversă este imposibilă, mai ales în cazul testelor de performanţă (Roşca, 1972).

Ambele moduri de examinare prezintă o serie de avantaje, dar şi limite. Examenul psihologic individual îi permite psihologului să stabilească un contact mai bun cu subiectul, să ajungă mai repede la cooperare cu acesta. Totuşi, decizia psihologului este mai încărcată de subiectivitate, iar examenul psihologic individual este mai puţin standardizat decât cel colectiv. În schimb, unele stări particulare ale subiecţilor, care pot denatura rezultatele testării, cum sunt oboseala, anxietatea, boala etc., sunt mai greu de detectat în condiţiile examenului colectiv.

Testarea individuală i se potriveşte mai bine psihologului clinician şi celui de consiliere în anumite activităţi de orientare şcolară şi profesională.

Testarea colectivă este preferată în selecţia profesională, când se operează cu grupuri mari de subiecţi, şi mai ales în examinările de tip screening.

B. După *timpul de execuţie* a sarcinii impuse, testele de aptitudini şi cele de cunoştinţe pot fi:

- *teste de viteză* (speed tests);
- *teste de randament* (power tests):
  - cu limită de timp;
  - cu timp de lucru nelimitat.

Testele de viteză conţin itemi la care orice persoană din categoria celor cărora le este destinată proba poate să răspundă cu uşurinţă corect, dar care trebuie rezolvaţi rapid. De exemplu, trebuie barate numere sau figuri sau trebuie identificate figurile identice. Ele sunt administrate cu limită de timp, fixată astfel încât nici o persoană să nu poată rezolva integral testul. Performanţa unui subiect este dată de numărul de itemi la care a răspuns. Aceste teste au la bază presupunerea că între viteza de lucru şi gradul de dezvoltare a caracteristicii măsurate de test există o asociere pozitivă.

Testele de randament conţin itemi de dificultăţi diferite, adesea aşezaţi în ordinea crescătoare a dificultăţii. Timpul de rezolvare a unui asemenea test nu este limitat sau este suficient de generos pentru a permite parcurgerea integrală a testului de către toate persoanele, iar performanţa unui subiect este indicată de numărul itemilor rezolvaţi corect.

Probele utilizate cel mai des sunt combinaţii ale celor două tipuri de teste. Ele conţin itemi de dificultăţi diferite şi sunt administrate cu limită de timp. Dacă timpul de lucru este atât de lung încât toate persoanele examinate au posibilitatea să rezolve toţi itemii la care cunosc răspunsul, proba se comportă ca un test de randament. Dacă timpul de lucru este prea scurt şi subiecţii nu au posibilitatea să rezolve toţi itemii la care ştiu răspunde, proba se comportă ca un test de viteză (Corsini, 1994).

Guilford (1965) apreciază că un test care a fost parcurs integral de cel puţin 75% dintre cei examinaţi poate fi considerat drept test de randament.

Atât testele de viteză, cât şi cele de randament sunt astfel construite încât să nu permită realizarea unui scor maxim (egal cu numărul itemilor testului). Acest lucru este necesar în vederea obţinerii unei imagini cât mai clare despre persoana testată, despre nivelul unei anumite aptitudini sau deprinderi a acesteia. Totodată, apare astfel posibilitatea unei ierarhizări mai precise a subiecţilor. În practică, distincţia între testele de viteză şi cele de randament este numai una de grad, cele mai multe teste depinzând de ambii factori în proporţii variabile (Cascio, 1991).

4.3. CLASIFICĂRI ÎN FUNCŢIE DE MODUL DE COTARE A TESTULUI ŞI DE INTERPRETARE A SCORURILOR

A. După *modul de stabilire a cotei la test*, se poate vorbi despre:

- *teste obiective*;
- *teste subiective.*

Un test obiectiv este cel al cărui scor se calculează urmând o regulă prestabilită, simplă, care cere o judecată redusă (English, English, 1970). De exemplu, se compară setul de răspunsuri date de subiect cu un set de răspunsuri fixat şi scorul subiectului la test este dat de numărul de răspunsuri care coincid. Termenul "obiectiv" se referă numai la modul în care se face cotarea răspunsurilor la test.

În cazul testelor obiective, comparativ cu cele subiective, persoanele examinate:

a. efectuează o sarcină complet structurată (ele nu pot să definească sau să organizeze problema de care se ocupă);

b. selectează un număr fixat de răspunsuri dintr-un număr fixat de posibilităţi de răspuns;

c. răspund la un număr mare de itemi;

d. primesc un scor la test calculat după o formulă stabilită aprioric (Thorndike, Hagen, 1961).

Pentru testele obiective scorul poate fi calculat de un operator (nu neapărat de psiholog) sau cu ajutorul unei maşini de cotare ori prin scanare şi cotare automată folosind calculatorul electronic.

Spre deosebire de testele obiective, testele subiective permit ca mai mulţi evaluatori să coteze foarte diferit răspunsurile aceleiaşi persoane. Din acest motiv astfel de teste sunt utilizate cu precădere de către psihologii clinicieni şi în unele cabinete de consiliere, mai puţin în mediul industrial.

Din categoria testelor subiective fac parte testele proiective.

B. După *modul în care se interpretează scorurile testelor*, se poate vorbi despre:

- *teste normative*;
- *teste centrate pe criteriu*;
- *teste idiografice*;
- *teste ipsative.*

- **Testele normative** (norm-referenced tests) interpretează un scor individual prin comparaţie cu scorurile obţinute la acelaşi test de un eşantion

reprezentativ de persoane. Se folosesc clasele etalonului (cuartile, decile, centile ori clase normalizate) sau valorile statistice (media şi abaterea standard) ale scorurilor în eşantionul reprezentativ. Cu ajutorul lor, pentru fiecare subiect se determină clasa din care face parte scorul obţinut sau se calculează cota transformată (cota z, cota T etc.).

Adesea, pentru un test se formează eşantioane reprezentative de persoane după mai multe criterii, de exemplu după vârstă, după profesie, după nivelul de şcolarizare etc. În acest caz, scorurile unui subiect pot fi comparate cu mai multe eşantioane de persoane, ceea ce permite o interpretare mai nuanţată.

• **Testele centrate pe criteriu** (criterion-referenced tests) sunt probe construite intenţionat pentru a furniza scoruri care să fie interpretabile direct, în termenii standardelor specificate pentru performanţa respectivă (R. L. Glaser şi A. J. Nitko, după Crehan, 1974).

Ele folosesc, pentru interpretarea scorurilor, un criteriu de performanţă, ales înainte de utilizarea testului. Pe baza acestui criteriu se stabileşte un "nivel acceptabil de performanţă" (Guilbert, 1987), cu care se compară scorul fiecărui individ. Rezultatul comparaţiei se exprimă de obicei dihotomic: subiectul îndeplineşte cerinţele criteriului (a realizat nivelul acceptabil de performanţă) sau nu le îndeplineşte.

Uneori, în urma analizei logice şi psihologice a sarcinilor impuse de test, se definesc mai multe clase de sarcini. Rezultatul obţinut de o persoană la test este comparat cu cerinţele fixate pentru fiecare clasă şi evaluarea subiectului constă în determinarea clasei de sarcini care îi este accesibilă (Petrovsky, Yaroshevsy, 1985).

Utilizarea testelor centrate pe criteriu pentru evaluarea elevilor are la bază ideea că prima preocupare a şcolii elementare nu este competitivă sau selectivă, ci formativă. Pentru a-l ajuta pe elev este mult mai puţin important să îi spui ce loc ocupă el în raport cu ceilalţi, conform cunoştinţelor sale, decât să îi arăţi cât de bogat este bagajul cunoştinţelor sale, cât de aproape se află de obiectivul urmărit de procesul de învăţământ (de Landsheere, 1975).

De exemplu, în cazul unui test de aritmetică format din 10 exerciţii de adunare, administrat la sfârşitul clasei întâi, se poate stabili drept nivel acceptabil de performanţă pentru promovarea clasei rezolvarea corectă a 6 exerciţii.

În scopul evitării confuziilor, termenul "criteriu" fiind folosit în evaluarea psihologică pentru a desemna variabila prezisă (şi corespunzător acestei accepţiuni se vorbeşte despre validitatea relativă la criteriu), American Psychological Association, American Educational Research Association şi National Council on Measurement au propus înlocuirea denumirii de "test

centrat pe criteriu" (criterion-referenced test) cu cea de "*test centrat pe conținut*" (content-referenced test) sau "*test centrat pe domeniu*" (domain-referenced test) (Silva, 1993). În literatura de specialitate, pentru acest tip de teste se mai folosește denumirea de "*test centrat pe obiective*" (de Landsheere, 1975).

Întrucât testele interpretate prin referire la criterii furnizează de obicei un scor numeric, ele pot fi utilizate și ca teste normative.

• **Testele idiografice** se utilizează în situațiile în care nu interesează compararea scorului unei persoane, la un chestionar, cu un criteriu sau cu rezultatele unui grup de persoane, ci evoluția în timp a scorului său (de exemplu, schimbările intervenite pe parcursul unui program psihoterapeutic). În asemenea cazuri, persoana completează o aceeași formă a chestionarului în mai multe ocazii și se urmăresc modificările apărute. Întrebările adresate unei persoane pot fi diferite de cele adresate alteia. Această abordare a testului se numește idiografică, deoarece se ocupă de un individ izolat, folosind un instrument construit numai pentru el.

Testarea idiografică are următoarele trei caracteristici (Silva, 1993):

a. Procesul de evaluare și instrumentul corespunzător sunt alese sau dezvoltate pentru fiecare individ.

b. Variabilele studiate provin de la subiect și din mediul său, ele nu sunt impuse din afară. Se ia în considerare percepția persoanei evaluate.

c. Evaluarea se ocupă cu descoperirea și explorarea patternului de variabile care este tipic pentru fiecare persoană și unic pentru aceasta.

Abordarea idiografică nu permite compararea între subiecți, dar este utilă pentru cercetarea proceselor dinamice care au loc în indivizi.

• **Testele ipsative** sunt cele la care interpretarea răspunsurilor date de un individ se face prin compararea între ele a scorurilor obținute de acesta la scalele sau la itemii testului (Corsini, 1994). De exemplu, testul *Edwards Personal Performance Schedule* măsoară 15 trebuințe. Prin compararea scorurilor realizate de un subiect la scalele probei se poate constata, de pildă, că pentru acesta nevoia de realizare este mai puternică decât nevoia de hrană.

În general, testele ipsative îi pretind subiectului să aloce o resursă proprie limitată (timp, bani etc.) mai multor domenii. Prin acest procedeu, atribuirea unei valori mari pentru un aspect scade automat valorile altor aspecte.

De exemplu, pentru studierea motivației sunt enumerate o serie de activități; persoana examinată trebuie să indice cât timp îi dedică fiecăreia, în mod obișnuit, într-o lună.

CAPITOLUL 5
CONSTRUIREA ŞI COTAREA ITEMILOR TESTELOR PSIHOLOGICE

5.1. TIPURI DE ITEMI ÎN TESTELE PSIHOLOGICE

Literatura de specialitate descrie şase tipuri de itemi de test utilizaţi frecvent în examinările psihologice. De obicei, aceştia sunt grupaţi în două categorii:

A - itemi cu răspunsuri închise - care solicită selectarea unuia sau a mai multor răspunsuri dintr-o listă de răspunsuri propuse :

- itemi dihotomici;
- itemi cu răspunsuri la alegere (cu alegere multiplă);
- itemi de tip asociere;

B - itemi cu răspunsuri deschise (libere) - care pretind formularea unui răspuns folosind vocabularul propriu:

- itemi cu formulări lacunare;
- itemi cu răspunsuri scurte;
- itemi-eseu.

Îi vom descrie pe scurt.

Itemii dihotomici pun în faţa subiectului fie o propoziţie - afirmaţie despre care acesta trebuie să spună dacă este "adevărată" ori "falsă" (de exemplu, itemii Inventarului Psihologic California), fie o întrebare directă, la care trebuie să răspundă prin "da" sau "nu" (de exemplu, itemii Chestionarului de personalitate Eysenck). Este vorba despre itemi cu conţinut dihotomic, cu numai două alegeri plauzibile.

Exemplu.

Faceţi zilnic gimnastică de înviorare? Răspundeţi prin "da" sau "nu".

Itemii dihotomici sunt uşor de elaborat, sunt cotaţi obiectiv şi se pot introduce în testele administrate cu ajutorul calculatorului. Răspunsurile la aceşti itemi pot fi prelucrate statistic fără nici o dificultate. Ca dezavantaj major este faptul că şansa de a ghici răspunsul corect, în cazul itemilor din testele de eficienţă (de aptitudini, de cunoştinţe) este mare, de 50%.

Identificarea răspunsului corect la itemii de tip "adevărat"/ "fals", în testele de cunoştinţe, este facilitată de multe ori de folosirea specificatorilor: cuvinte ca "totdeauna", "pretutindeni", "niciodată" sau "numai" sugerează că

răspunsul corect este "fals", în timp ce cuvinte ca "adesea", "uneori" sau "obişnuit" conduc spre răspunsul "adevărat" (Aiken, 1994).

A. L. Edwards (Gregory, 1992) susţine că răspunsurile date de subiecţi la itemi de acest fel, în chestionarele de personalitate, reflectă mai mult dezirabilitatea socială decât trăsătura de personalitate. Pentru a se evita aceasta, se pot utiliza itemi cu alegere forţată, la care subiectul să fie obligat să aleagă un răspuns din două propuse, ambele având acelaşi grad de dezirabilitate socială.

Exemplu.

În vacanţe îmi petrec timpul mai ales:
a. în faţa televizorului;
b. făcând sport.

Alegeţi răspunsul care vi se potriveşte.

Pentru a nu verifica doar memorarea unor cunoştinţe, ci şi înţelegerea lor, itemii dihotomici pot întreba dacă:

- un eveniment este cauză (sau efect) al altuia;
- un obiect, un eveniment sau o condiţie face parte dintr-o anumită categorie;
- două evenimente pot avea loc simultan.

Itemii cu răspunsuri la alegere conţin o întrebare şi o listă de răspunsuri propuse sau o frază incompletă şi o listă de completări posibile. Subiectului i se cere să selecteze din listă răspunsurile sau completările care i se potrivesc (în cazul testelor de personalitate) sau pe care le consideră corecte (în cazul testelor de eficienţă). Aşa sunt, de exemplu, itemii Chestionarului 16PF al lui Cattell sau cei din testul Matrici Progresive Raven.

Alteori, itemul are forma unui scenariu complex care poate conţine câteva probleme, la care trebuie să se dea mai multe răspunsuri prin selectarea celor corecte dintr-o mulţime de răspunsuri propuse.

Exemplu.

Aurel, Barbu, Călin şi Dan au participat la un concurs de fugă. După stabilirea clasamentului, s-au constatat următoarele:

- Dan nu a ocupat nici primul, nici ultimul loc.
- Călin s-a clasat în urma lui Aurel.
- Barbu a obţinut un timp mai bun decât Dan.
- Între Aurel şi Barbu s-a clasat o altă persoană.

1. Cine s-a clasat pe locul al doilea?
2. Dar pe ultimul?

Răspundeţi la fiecare întrebare cu litera aflată în faţa numelui persoanei, în lista următoare:

a. Aurel;
b. Barbu;
c. Călin;
d. Dan. Răspunsuri: 1:.... 2:....

Sub formă de itemi cu alegere multiplă pot fi redactate şi întrebări complexe. Aiken (1994) propune ca prin itemi de această formă să li se ceară subiecţilor să efectueze următoarele sarcini:

A. să claseze o persoană, un obiect sau o condiţie în una din categoriile prezentate în lista de răspunsuri propuse;

Exemplu.

Metoda formelor echivalente se utilizează în studiul:
a. validităţii predictive;
b. validităţii concurente;
c. fidelităţii;
d. puterii de discriminare.

B. să decidă care este consecinţa corectă a unuia sau a mai multor evenimente prezentate în trunchiul itemului;

Exemplu.

Dacă se înlocuieşte un item al unui test cu un alt item, care are aceeaşi dispersie a scorurilor ca şi primul, dar corelează mai puternic decât acesta cu scorul total al testului, atunci fidelitatea testului:
a. va scădea;
b. va creşte;
c. va rămâne neschimbată;
d. se va schimba, dar nu se poate preciza cum.

C. să utilizeze două sau mai multe condiţii, prezentate în trunchiul itemului, pentru a trage o concluzie;

Exemplu.

Dacă unui test format din 20 de itemi paraleli, care are coeficientul de fidelitate egal cu 0.6, i se adaugă 5 itemi paraleli, ce valoare va avea coeficientul de fidelitate al noului test?
a. 0.45;
b. 0.65;
c. 0.75;
d. nici unul dintre aceştia.

D. să decidă dacă dintre două afirmaţii cuprinse în trunchiul itemului este adevărată numai una sau ambele sunt adevărate sau ambele sunt false;
Exemplu.

Se dau următoarele afirmaţii:
(1) Coeficientul ρ_{20} serveşte la calculul coeficientului de fidelitate prin metoda formelor echivalente;
(2) Coeficentul ρ_{20} se calculează numai pentru teste formate din itemi dihotomici.

Care este valoarea de adevăr a acestor afirmaţii?
a. ambele sunt adevărate;
b. este adevărată numai prima;
c. este adevărată numai a doua;
d. ambele sunt false.

E. să determine o opţiune care nu îşi are locul printre celelalte;
Exemplu.
Care dintre următorii coeficienţi **nu** face parte din aceeaşi categorie cu ceilalţi:
a. ρ_{20};
b. ρ_{21};
c. α al lui Cronbach;
d. λ_3 al lui Gutman;
e. coeficientul lui Spearman-Brown.

F. să determine relaţia dintre două noţiuni (1 şi 2) şi să identifice noţiunea, din lista de răspunsuri propuse, care este legată de noţiunea 3, aşa cum sunt legate între ele noţiunile 1 şi 2.
Exemplu.

Numărul itemilor paraleli din test şi coeficientul lui Spearman-Brown sunt la fel ca şi numărul scorurilor posibile ale testului şi:
a. dificultatea testului;
b. puterea de discriminare a testului;
c. validitatea testului;
d. fidelitatea test-retest.

Itemii cu răspunsuri la alegere sunt foarte frecvent utilizaţi în testele de cunoştinţe administrate creion-hârtie sau cu ajutorul calculatorului. Sunt cotaţi obiectiv. Probabilitatea de ghicire a răspunsului corect este, în general, mai mică decât în cazul itemilor dihotomici, ceea ce constituie un avantaj. Răspunsurile la aceşti itemi se pretează prelucrărilor statistice.

Ca dezavantaj, amintim consumul mare de timp necesar elaborării unor itemi în care răspunsurile propuse să fie toate la fel de atractive pentru subiecţii care nu cunosc răspunsul corect.

Itemii de tip asociere constau din două liste de elemente şi o regulă de asociere a câte două elemente, unul dintr-o listă şi celălalt din a doua. Subiectului i se cere să alcătuiască o listă-răspuns, prin formarea de perechi. Obişnuit, itemii discutaţi se prezinta ca o listă de cuvinte care trebuie alocate unei alte liste, care conţine definiţii.

Exemplu.

Pentru fiecare parametru al funcţiei caracteristice a unui item indicaţi denumirea sa:

Parametru	Denumire
a	1. puterea de discriminare
b	2. parametru de ghicire
c	3. dificultatea itemului
	4. fidelitatea itemului

Este recomandabil ca listele să nu conţină acelaşi număr de elemente.

Itemii de tip asociere sunt cotaţi obiectiv. Ei pot fi introduşi în testele administrate cu ajutorul calculatorului.

O deficienţă a acestor itemi constă în faptul că, dacă fiecare element al unei liste poate face parte dintr-o singură pereche, atunci răspunsurile date de o persoană nu sunt independente între ele: introducerea unui element într-o pereche nu mai dă posibilitatea utilizării sale şi în altă pereche. Prin urmare, formarea unei perechi greşite va avea drept consecinţă apariţia altor perechi eronate. Se poate evita acest fapt permiţând formarea mai multor perechi cu elementele unei liste.

Itemii cu formulări lacunare sunt de format deschis, adică pretind construirea răspunsului şi nu alegerea acestuia dintr-o listă. Ei adresează o întrebare la care se cere să se formuleze un răspuns sau prezintă un text conţinând spaţii libere, care trebuiesc completate.

Exemple.

1. Care este valoarea minimă a coeficientului de fidelitate al unui test psihologic?

2. Proprietatea unui test de a-şi îndeplini funcţia pentru care a fost construit se numeşte ...

Completaţi denumirea.

Itemii cu formulări lacunare pot fi introduşi în testele administrate de calculator. Dar, trebuie să se prevadă toate posibilităţile de a se răspunde corect (de exemplu, toate sinonimele răspunsului aşteptat).

Principalul dezavantaj prezentat de aceşti itemi este faptul că nu se poate asigura totdeauna obiectivitatea cotării: uneori sunt formulate răspunsuri diferite de cele aşteptate, iar cotarea acestora este stabilită de evaluator.

Itemii cu răspunsuri scurte îi cer persoanei testate să formuleze un răspuns clar pe cel mult o pagină.

Aceşti itemi sunt recomandaţi pentru a fi utilizaţi în testele de cunoştinţe, când se urmăreşte în ce măsură subiectul a înţeles materialul învăţat, este capabil să explice, să analizeze sau să sintetizeze. De asemenea, ei sunt utili atunci când se doreşte evaluarea creativităţii.

Exemplu.

Care dintre formele de validitate este mai importantă în cazul unui test utilizat în selecţia profesională ? Argumentaţi răspunsul.

Avantajul major al itemilor de acest tip este faptul că permit aprecierea originalităţii răspunsului.

Principalul dezavantaj îl constituie lipsa obiectivităţii în cotare. Scorul acordat unui asemenea item este influenţat de efectul de halo, de oboseala celui care face notarea, de calitatea redactării etc.

Itemii eseu pretind răspunsuri cu format deschis, a căror lungime depăşeşte o pagină.

Exemplu.

Sunteţi psiholog la o firmă care vrea să angajeze un programator. Vi se cere să realizaţi examinarea psihologică în acest scop. Cum procedaţi ?

Aceşti itemi sunt recomandaţi atunci când interesează capacitatea subiectului de a formula un răspuns original.

Ei permit să se măsoare capacitatea persoanelor de a organiza, de a relata sau/ şi de a comunica, adică ceea ce itemii cu răspuns închis nu pot evalua.

La un asemenea item, într-un test de cunoştinţe, subiecţii nu pot ghici răspunsul corect, ceea ce constituie un avantaj, dar pot evita să discute despre ceea ce cunosc mai puţin, dezvoltând subiectele pe care le stăpânesc.

Itemii eseu sunt uşor de construit, dar consumă mult timp pentru evaluarea răspunsului şi nu pot fi cotaţi obiectiv.

Dacă nu pot fi înlocuiţi cu itemi cu răspunsuri închise, pentru a se micşora subiectivitatea cotării, itemii de tip eseu vor fi formulaţi astfel încât sarcina subiecţilor să fie bine precizată: li se va cere să "explice" şi să "stabilească deosebiri", nu să "discute".

În chestionarele de autoevaluare apar câteva tipuri de itemi asemănătoare ca formă itemilor cu răspunsuri la alegere. Aşa sunt:

- scalele Likert;
- itemii cu alegere forţată;
- scalele cu ancore comportamentale.

Îi vom descrie în continuare.

O **scală Likert** constă dintr-o afirmaţie însoţită de mai multe răspunsuri propuse (de obicei 3-7) care reprezintă grade diferite de acord cu ideea exprimată în enunţ. De exemplu, asemenea răspunsuri pot fi:

5 - îmi place totdeauna;
4 - îmi place adesea;
3 - uneori îmi place, alteori nu;
2 - adesea nu-mi place;
1 - niciodată nu-mi place.

Subiectul trebuie să indice răspunsul care i se potriveşte cel mai mult.

O scală Likert se comportă ca un item. Ea este cotată independent de răspunsurile date la alte scale. De obicei cota este numărul de ordine al răspunsului.

Avantajul prezentat de scalele Likert faţă de itemii dihotomici constă în faptul că ele permit o mai mare varietate de răspunsuri şi în acest fel se diminuează riscul de a obţine de la majoritatea persoanelor acelaşi răspuns, cel care corespunde dezirabilităţii sociale.

Un **item cu alegere forţată** conţine două, trei sau mai multe caracteristici, activităţi ori alte descrieri care au legătură cu ceea ce măsoară testul, toate la fel de atractive. Subiectul trebuie să aleagă ceea ce îi place cel mai mult şi ceea ce îi place cel mai puţin sau ceea ce consideră că este cel mai descriptiv şi cel mai puţin descriptiv. De exemplu, un asemenea item poate cere să se indice activitatea cea mai plăcută dintre "ascultarea unui concert de muzică uşoară" şi "vizionarea unui film".

Itemii cu alegere forţată corespund situaţiilor din viaţa zilnică, întrucât în mod obişnuit nu se întâmplă ca mai multe lucruri să fie la fel de atractive în acelaşi timp.

O scală de autoevaluare cu ancore comportamentale se referă la o dimensiune a constructului măsurat de test. Se prezintă sub forma unui segment gradat alături de care sunt fixate o serie de *ancore* care reflectă niveluri diferite de dezvoltare a dimensiunii respective. Fiecare ancoră conţine o descriere a unui comportament, apreciat ca fiind caracteristic unei persoane care se află la nivelul aflat în dreptul ancorei.

Subiectul trebuie să marcheze pe segmentul scalei gradaţia care crede că i se potriveşte cel mai mult.

Numărul de gradaţii variază de obicei între 4 şi 8. Numărul de ancore nu trebuie să fie egal cu numărul de gradaţii. Ele nu sunt repartizate echidistant de-a lungul segmentului ci sunt plasate în dreptul nivelurilor la care se potrivesc.

Exemplu.

Scala "Controlul termenelor" din *Chestionarul de Percepţie a Timpului* (Landy, Rastegary, Thayer şi Colvin, 1991), conţine următoarele ancore, ordonate de la nivelul cel mai scăzut de manifestare a acestei dimensiuni (cotat cu 1) spre nivelul cel mai ridicat (cotat cu 7):

- Adesea solicit derogări de la termenele de execuţie sau de la diferite sarcini.
- Rar îmi stabilesc termene de execuţie.
- Prefer să mă concentrez asupra calităţii muncii mele şi nu asupra efectuării ei la timp sau după un grafic/termen scadent.
- Îmi stabilesc termene de execuţie pentru mine însumi, doar dacă este necesar.
- Uneori sunt preocupat de termenele de execuţie ale unei activităţi.
- Sunt totdeauna pregătit pentru orice eveniment.
- Dacă nu îmi este fixat un termen de execuţie pentru o sarcină, îl stabilesc eu însumi.
- Îmi achit notele de plată (taxele de lumină, telefonul etc.) cât pot de repede.

Întrucât o asemenea scală este cotată independent de altele, ea se comportă ca un item. Dar, pentru că scorul său nu se adună la scorurile altora şi este interpretat separat, ea este considerată ca fiind o scală compusă dintr-un singur item.

5.2. SUGESTII PENTRU SCRIEREA ITEMILOR

Indiferent de tipul itemului, la redactarea sa este bine să se aibă în vedere următoarele reguli (după Ebel, 1955; de Landsheere, 1975; Shrock, Coscarelli, 1989):

a. Formularea itemului să fie adecvată nivelului de maturitate şi de pregătire al persoanelor cărora le este destinat testul.

b. Ideile să fie exprimate clar, prin cuvintele cele mai potrivite, cu un sens precis. Se vor evita frazele lungi şi asocierile de cuvinte care fac dificilă înţelegerea întrebării.

c. Să nu se introducă în item informaţii neesenţiale sau cuvinte nefuncţionale (care nu contribuie la fundamentarea alegerii răspunsului).

d. Să fie incluse în întrebare toate datele necesare pentru a permite raţionamentul pe baza căruia se alege răspunsul. Nu se vor considera implicite anumite informaţii. S-ar putea ca nu toţi subiecţii să le cunoască sau să le presupună.

e. Să se evite formulările care sugerează răspunsul.

f. Structura textului întrebării să fie cât mai simplă. În general este preferabil ca itemul să înceapă cu întrebarea, iar răspunsurile propuse şi explicaţiile să fie aşezate la urmă.

g. Să se evite utilizarea, în textul întrebării, a negaţiilor (cum sunt "nu", "nici", "niciodată"), în special a negaţiilor duble. Dacă, totuşi, negaţiile sunt necesare, ele vor fi scrise cu alte caractere (de exemplu, cu majuscule) sau vor fi subliniate pentru a atrage atenţia asupra lor.

Următoarele sugestii se referă la itemii testelor de eficienţă.

Itemii de tipul "adevarat-fals" trebuie să se bazeze numai pe declaraţii care sunt absolut şi fără ambiguitate adevărate sau false, adică a căror valoare de adevăr este apreciată la fel de către orice persoană care are aptitudinea, deprinderea sau cunoştinţele necesare pentru a da un răspuns corect.

Dacă declaraţiile sunt numai aproximativ adevărate, atunci în faţa celui examinat se ridică o problemă dificilă: el trebuie să ghicească ce toleranţă este admisă la aprecierea corectitudinii răspunsului. De exemplu, o aceeaşi întrebare de tip "adevărat-fals" poate fi formulată în urmatoarele trei moduri:

1. Valoarea lui π este 3,14.
2. Valoarea lui π este 3,1416.
3. Valoarea lui π cu 4 zecimale este 3,1416.

Primele două variante sunt doar parţial adevărate. Ele vor fi considerate false nu numai de către subiecţii care nu cunosc valoarea numarului π, ci şi de cei foarte bine informaţi, care ştiu că 3.14 şi, respectiv, 3.1416 sunt numai aproximaţii ale acestuia.

În test, itemii pentru care răspunsul corect este "fals" să fie aproximativ de aceeaşi lungime şi în acelaşi număr ca şi itemii al căror răspuns corect este "adevărat".

Pentru **întrebările cu alegere multiplă** este bine să se ia în considerare urmă-toarele recomandări (Ebel, 1955; Shrock şi Coscarelli, 1989; Gregory, 1992):

a. Numărul de răspunsuri propuse să fie cuprins între 3 şi 5.

b. Răspunsurile propuse să aibă aproximativ aceeaşi lungime.

c. Răspunsurile propuse vor fi aşezate în ordine logică. De exemplu, daca acestea sunt numere, ele vor fi aşezate în ordine crescătoare sau descrescătoare.

d. Se va urmări ca răspunsurile propuse să nu se suprapună şi să nu se includă unul pe altul.

e. Se va evita plasarea răspunsului corect pe aceeaşi poziţie într-un număr mare de întrebări.

f. Ca răspunsuri incorecte, în lista de răspunsuri propuse, se vor utiliza erorile pe care le fac în mod obişnuit subiecţii din categoria celor cărora li se adresează testul.

g. Trunchiul itemului va conţine toate informaţiile necesare pentru alegerea răspunsului corect.

h. Pentru simplificarea formulării, se vor include în trunchiul itemului (care conţine fie o întrebare directă, fie o declaraţie incompletă) cuvintele care ar trebui repetate în fiecare răspuns, dacă acest lucru nu îngreunează înţelegerea.

i. Se va respecta acordul gramatical între textul inclus în trunchiul itemului şi răspunsurile propuse.

j. Se va evita utilizarea unor declaraţii negative în trunchiul itemului. Dacă, însă, este dificil de reformulat itemul pentru a se elimina declaraţia negativă, cuvântul "nu" din textul întrebării trebuie subliniat sau scris cu majuscule.

k. Expresiile "nici una dintre acestea" şi "toate acestea" vor fi utilizate ca răspuns propus numai atunci când se poate da un răspuns absolut corect la întrebare. Se recomandă ca aceste formulări să fie folosite ca răspuns incorect.

l. În acelaşi test pot să apară itemi cu un număr diferit de răspunsuri propuse.

Pentru scrierea **itemilor de tip asociere** sunt utile următoarele sugestii (Shrock, Coscarelli, 1989):

a. Elementele fiecărei liste să fie apropiate între ele, să aparţină aceleiaşi categorii.

b. Fiecare listă poate conţine între 5 şi 15 elemente, aranjate, dacă este posibil, într-o ordine logică (cronologic, crescător sau descrescător după valoarea numerică etc.).

c. În trunchiul întrebării se va prezenta regula de formare a perechilor.

d. Se va specifica dacă un element al unei liste poate fi inclus într-o singură pereche sau în mai multe.

e. Cel puţin una dintre liste va conţine şi elemente care nu pot fi cuprinse în nici o pereche.

Utilizarea, în teste, a **întrebarilor cu răspunsuri deschise** ridică probleme atât în ceea ce priveşte formularea întrebărilor, cât şi la interpretarea şi punctarea răspunsurilor. Experienţa a dovedit că este aproape imposibil să se formuleze itemii astfel încât toţi cei care cunosc răspunsul să îl exprime identic. Autorii de itemi au fost surprinşi de varietatea mare a răspunsurilor corecte care apar la întrebări pentru care s-a prevăzut un singur răspuns.

Se întâmplă, uneori, ca răspunsul dat de subiecţi să fie corect, dar îndepărtat de intenţia autorului itemului. De exemplu, la întrebarea "Poetul G. Coşbuc s-a născut în ..." se poate răspunde corect fie cu "1866", fie cu "Hordou".

Este necesară, deci, o formulare atentă, care să furnizeze elemente suficiente pentru a-l conduce pe cel examinat spre răspunsul aşteptat. Astfel, în exemplul anterior, textul itemului ar fi trebuit completat cu unul dintre cuvintele "localitatea" sau "anul".

Chiar şi problemele matematice pot cauza asemenea necazuri, dacă nu se specifică precizia rezultatului cerut, forma de exprimare (printr-o fracţie ori printr-un număr zecimal) sau unitatea de măsură.

Este posibil ca un răspuns corect să admită mai multe sinonime. Întrucât aprecierea corectitudinii unui răspuns la o întrebare deschisă se face prin compararea răspunsului dat de subiect cu cel indicat de către autorul itemului este necesar să se prevadă toate variantele corecte posibile.

În cazul întrebărilor cu răspunsuri deschise trebuie să i se specifice subiectului dacă la cotare se ţine seama şi de regulile ortografice şi gramaticale.

De asemenea, trebuie să i se indice cât spaţiu are la dispoziţie pentru redactarea răspunsului (număr de rânduri, pagini sau ecrane).

5.3. CALCULUL SCORULUI UNUI ITEM

Calculul scorului unei persoane la un test presupune, în general, parcurgerea următorului algoritm:

(1) Se calculează punctajul pentru fiecare item al testului.
(2) Se înmulţeşte scorul fiecărui item cu câte o pondere.
(3) Se însumează valorile obţinute la pasul 2.

Întrucât valorile ponderilor depind numai de caracteristicile itemilor, ele pot fi incluse în scorurile acestora. În acest caz, pasul 2 nu mai are rost.

În literatura de specialitate (Cronbach, 1966; Fisseni, 1990; Guilford, 1936 şi 1965; Lord, Novick, 1968; Traxler, 1955) sunt prezentate diverse formule de calcul al scorului itemilor.

A. Pentru itemii din **chestionarele de personalitate** procedeele utilizate în majoritatea cazurilor sunt următoarele:
a. Itemii dihotomici primesc un punct dacă răspunsul denotă un nivel mai înalt al caracteristicii măsurate de test şi zero dacă indică un nivel mai scăzut.
b. Itemii cu alegere multiplă au asociată câte o cotă pentru fiecare răspuns propus. Ordinea răspunsurilor stabilită pe baza acestor cote trebuie să corespundă nivelurilor succesive ale variabilei măsurate de test: cota cea mai mare o primeşte răspunsul care este ales de persoanele cu cel mai înalt nivel al acestei variabile.
c. Pentru scalele Likert, cota este egală de obicei cu numărul de ordine al răspunsului în şirul ordonat în care primul răspuns corespunde nivelului cel mai scăzut al variabilei măsurate de test.

B. Scalele de autoevaluare comportamentală bazate pe ancore au cota egală cu gradaţia indicată de subiect.

C. Pentru **testele de aptitudini** şi **testele de cunoştinţe**, atunci când răspunsul la item este apreciat prin "corect" sau "greşit", se pot aplica diverse formule de calcul al scorului. Vom prezenta în continuare câteva dintre ele.

Cea mai simplă tehnică de cotare a itemilor unui test de eficienţă, valabilă pentru toate tipurile de întrebări, constă în acordarea unui punct fiecărui răspuns corect şi zero puncte răspunsurilor greşite sau omise. În felul acesta, la un item care pretinde k răspunsuri (de exemplu, selectarea a k răspunsuri dintr-o listă de răspunsuri propuse sau enumerarea a k termeni etc.) se poate obţine un scor cuprins între zero puncte (dacă nu s-a dat nici un răspuns corect) şi k puncte (dacă s-au indicat corect toate răspunsurile cerute). Deci punctajul maxim posibil este dat de numărul răspunsurilor corecte existente la întrebare, neavând nici o importanţă numărul răspunsurilor posibile. Or, cu cât sunt mai numeroase răspunsurile posibile, cu atât este mai dificil să se selecteze răspunsurile corecte. La o întrebare de tip "adevărat-fals", probabilitatea de a obţine un punct, ghicind răspunsul, este de 50%, în timp ce şansa de a ghici răspunsul corect din 5 răspunsuri propuse, bine formulate -încât toate să fie plauzibile- este de 20%; iar probabilitatea de a ghici anul naşterii lui Mihai Eminescu, la o întrebare deschisă care solicită acest lucru, este mult mai mică. O soluţie pentru a evita

obţinerea aceluiaşi punctaj la itemi care măsoară aceeaşi caracteristică, dar care, datorită formei de prezentare sunt de dificultăţi diferite, ar fi să se utilizeze metoda de calcul al scorului expusă mai sus numai în cazul testelor alcătuite din itemi care au aceeaşi probabilitate de ghicire a răspunsului corect.

Un alt dezavantaj al acestui mod de punctare constă în faptul că permite obţinerea scorului maxim la o întrebare cu alegere multiplă şi atunci când nu se cunosc răspunsurile corecte, dar se selectează toate răspunsurile propuse. În orice caz, sunt avantajaţi cei care, nefiind siguri pe cunoştinţele lor, aleg cât mai multe răspunsuri propuse.

Chiar şi atunci când se specifică numărul răspunsurilor aşteptate, această metodă îi favorizează pe cei care încearcă să ghicească răspunsurile ori de câte ori nu le cunosc.

De exemplu, în cazul unei întrebări dintr-un test de cunoştinţe, care solicită selectarea a *k* răspunsuri din *n* propuse ($2k \leq n$), o persoană care nu cunoaşte deloc materia verificată, alegând la întâmplare *k* răspunsuri, are o şansă de

$$100\ x \left(1 - \frac{C_{n-k}^{k}}{C_{n}^{k}}\right) \qquad (5.3.1)$$

la sută să obţină cel puţin un punct. (Dacă n=4 şi k=2, atunci probabilitatea de a ghici cel puţin un răspuns corect este de 83%.) Când $n < 2k$, adică mai mult de jumătate din răspunsurile propuse sunt corecte, atunci orice alegere de *k* itemi aduce un minimum de $2k$-n puncte. În asemenea situaţii, întrebarea ar trebui reformulată, astfel încât să ceară selectarea celorlalte n-k răspunsuri din cele *n* propuse.

Metoda este utilă, însă, în cazul întrebărilor deschise, atunci când probabilitatea de a ghici răspunsurile corecte este scăzută.

În cazul testelor compuse din întrebări cu alegere multiplă, dacă fiecare item cere selectarea unui singur răspuns din *n* propuse, pentru calculul scorului total al testului se poate utliza următoarea formulă, cunoscută sub denumirea de **formula corectată pentru şansa succesului:**

$$S = C - \frac{G}{n-1} \qquad (5.3.2)$$

unde:

S = scorul testului;
C = numărul itemilor la care s-a răspuns corect;
G = numărul itemilor la care s-a răspuns greşit.

Ea se bazează pe ipoteza că orice persoană care nu cunoaşte răspunsul la un item îl omite sau ghiceşte la întâmplare. Dar, această presupunere este rareori corectă. Examinaţii care au informaţii parţiale despre un item nu vor răspunde la întâmplare. La fel, persoanele care au informaţii greşite. În asemenea situaţii, răspunsurile greşite propuse la întrebare nu sunt la fel de atractive pentru examinat.

Adevărul este că persoanele diferă mult între ele în ceea ce priveşte modul de a proceda faţă de itemii al căror răspuns nu îl cunosc. Unele nu răspund, altele răspund la întâmplare.

O proprietate a formulei corectate pentru şansa succesului este că, dacă pentru un item toţi examinaţii ghicesc la întâmplare, atunci scorul mediu la întrebare este egal cu zero. Dacă, însă, persoanele testate sunt capabile să elimine corect y răspunsuri dintre cele n propuse şi vor alege la întâmplare dintre cele n-y răspunsuri rămase, atunci media scorurilor la item va fi

$$m = \frac{1 - \dfrac{n-y-1}{n-1}}{n-y} = \frac{y}{(n-y)(n-1)} \qquad (5.3.3)$$

deci mai mare decât zero. Dacă răspunsurile propuse la o întrebare nu sunt atât de bine formulate încât cele greşite să fie mai atractive decât răspunsul corect, persoanele care nu cunosc răspunsul corect, dar aleg inteligent, vor fi avantajate faţă de cei care nu răspund.

Ori de câte ori o persoană poate elimina cel puţin un răspuns greşit din cele propuse, va avea şanse să obţină un punctaj mai mare alegând la întâmplare unul dintre răspunsurile propuse rămase. Şi, de fapt, este normal ca un examinat care are informaţii parţiale despre un item să obţină un scor mai mare la test decât persoanele care nu au deloc cunoştinţele verificate de respectiva întrebare.

O practică larg acceptată în administrarea testelor este de a le spune examinaţilor că este în dezavantajul lor să ghicească răspunsul atunci când nu îl cunosc, deoarece greşelile sunt penalizate la calculul scorului, dar de a-i încuraja să răspundă de fiecare dată când nu sunt siguri, însă pot face un raţionament bazat pe informaţii, asupra celui mai rezonabil răspuns. (Cronbach, 1966). Un asemenea procedeu scade numărul omisiunilor.

În faza analizei de itemi constructorul de test are interesul să obţină cât mai multe răspunsuri, pentru a putea identifica răspunsurile propuse mai puţin

plauzibile, spre a le reformula sau elimina. În această situaţie, este recomandabil să se aplice modul de lucru descris mai sus.

Dacă nu există omisiuni, scorurile calculate cu formula corectată pentru şansa succesului sunt corelate liniar cu scorurile obţinute atribuind câte un punct pentru fiecare răspuns corect. Îndiferent cu care dintre formule s-ar calcula scorurile, ordinea persoanelor examinate, stabilită pe baza scorurilor, este aceeaşi.

O metodă de calcul al scorului pentru un test compus din itemi cu alegere multiplă, mai generală decât formula corectată pentru şansa succesului, a fost propusă de Traxler (1955). Ea se poate utiliza atunci când toţi itemii testului propun acelaşi număr de răspunsuri, *n*, şi cer selectarea aceluiaşi număr de răspunsuri corecte, *k*:

$$S = C - \frac{k}{n-k}G \qquad (5.3.4)$$

unde:
S = scorul testului;
C = numărul de răspunsuri corecte date de subiect;
G = numărul de răspunsuri greşite date de subiect.
Pentru *k*=1 se regăseşte formula corectată pentru şansa succesului. Se poate obţine scorul *S*, dat de formula de mai sus, dacă se acordă:
1 punct pentru fiecare răspuns corect;
- *k*/(*n*-*k*) puncte pentru fiecare răspuns greşit;
0 puncte pentru fiecare răspuns omis.
Înseamnă că scorul la un item va fi cuprins între -*k* puncte (dacă se dau *n*-*k* răspunsuri, toate greşite) şi *k* puncte (dacă se dau *k* răspunsuri corecte). Ori de câte ori nu se răspunde la un item sau se indică toate răspunsurile propuse, scorul va fi nul.

Cantitatea scăzută pentru un răspuns greşit, *k*/(*n*-*k*), este cu atât mai mare, cu cât raportul *k*/*n*, adică probabilitatea unui răspuns corect, este mai mare.

Avantajul acestei formule constă în faptul că permite punctarea şi pentru itemii rezolvaţi parţial corect. Dar, formula are şi o deficienţă majoră: dacă se modifică o întrebare astfel încât să ceară selectarea celor *n*-*k* răspunsuri care în prima variantă erau considerate greşite, atunci intervalul de valori pentru scor se schimbă din [-*k*, *k*] în [-(*n*-*k*), (*n*-*k*)], cu toate că, itemul măsoară acelaşi lucru ca şi înainte.

Pentru remedierea deficienţei semnalate, formula de mai sus ar putea fi modificată astfel:

$$S = \frac{C}{k} - \frac{G}{n-k} \quad (5.3.5)$$

Atunci intervalul de valori pentru scorul oricărui item devine [-1, 1].

Itemii de tip asociere pot fi consideraţi, pentru calculul scorului, drept itemi cu alegere multiplă, în care numărul de răspunsuri propuse, *n*, este egal cu numărul tuturor perechilor care pot fi formate, iar numărul de răspunsuri corecte, *k*, este egal cu numărul perechilor corecte.

De exemplu, să presupunem că întrebarea propune următoarele două liste:

A1	B1
A2	B2
	B3

şi cere să se formeze trei perechi în care un element să fie dintr-o listă, iar celălalt, din a doua. Regula după care se alcătuiesc aceste perechi este formulată în textul întrebării. Să presupunem că sunt corecte perechile (A1,B1), (A1,B2) şi (A2,B1).

Se interpretează întrebarea ca fiind un item cu alegere multiplă, care propune 2x3=6 răspunsuri: (A1,B1), (A2,B1), (A1,B2), (A2,B2), (A1,B3), (A2,B3). În acest caz, *n*=6 şi *k*=3.

CAPITOLUL 6
PRINCIPALELE TEORII ASUPRA TESTELOR PSIHOLOGICE

6.1. TEORIA CLASICĂ A TESTELOR

6.1.1. Scor real, scor observat

Se consideră două mulţimi :
P = o mulţime de persoane;
M = o mulţime de instrumente de măsurăi(teste sau itemi de test).

Ambele mulţimi pot fi infinite - dar numărabile- sau pot conţine un număr finit de elemente; în particular, ele pot fi formate dintr-un singur element.

Rezultatul efectuării unei măsurători cu un instrument $m \in \mathbf{M}$ asupra unei persoane $p \in \mathbf{P}$ se exprimă sub forma unui număr real, care se notează cu x_{mp}, şi se numeşte **scor observat**. Valoarea lui depinde de condiţiile particulare în care se face măsurarea. x_{mp} este doar una dintre valorile care s-ar putea obţine pentru instrumentul de măsură $m \in \mathbf{M}$ şi persoana $p \in \mathbf{P}$.

Raţionând în acest mod, putem interpreta valoarea x_{mp} ca fiind o realizare a unei variabile aleatoare pe care o notăm cu X_{mp}. Repartiţia variabilei aleatoare X_{mp}, dată de toate scorurile observate care s-ar putea obţine prin efectuarea de măsurători, în diverse condiţii, folosind instrumentul m asupra persoanei p, cu probabilităţile corespunzătoare, nu este cunoscută. În primul rând, este imposibil chiar şi numai de imaginat, cu atât mai mult de realizat practic, toate situaţiile particulare în care ar putea fi efectuată măsurarea. Pe de altă parte, două scoruri observate nu sunt întotdeauna rezultatul unor experimente independente. O asemenea situaţie poate să apară, de exemplu, atunci când răspunsurile date la o administrare a unui test sunt memorate de subiect şi reproduse mecanic la administrările ulterioare.

Pentru variabila aleatoare X_{mp} vom nota valoarea medie cu τ_{mp}:

$$\tau_{mp} = \mathrm{M}(X_{mp}) \qquad (6.1.1.1)$$

Întrucât repartiţia variabilei aleatoare X_{mp} nu este cunoscută, valoarea medie τ_{mp} nu poate fi calculată în practică. Dar, teoretic, ea există. Vom presupune că ea este finită, adică $\tau_{mp} < \infty$.

Pentru un instrument de măsură $m \in \mathbf{M}$ şi o persoană $p \in \mathbf{P}$ numărul real τ_{mp} se numeşte **scor real**. El poate fi interpretat ca valoare unică a unei variabile aleatoare constante T_{mp}. Guilford (1965) apreciază că scorul real este valoarea

care s-ar obţine dacă măsurarea ar fi făcută cu un instrument perfect, aplicat în condiţii ideale.

Diferenţa dintre un scor observat, x_{mp}, şi scorul real τ_{mp} se numeşte în mod obişnuit **eroare de măsură** şi se notează cu e_{mp}:

$$e_{mp} = x_{mp} - \tau_{mp} \quad (6.1.1.2)$$

e_{mp} poate fi considerat drept o valoare posibilă a variabilei aleatoare E_{mp} definită prin relaţia:

$$E_{mp} = X_{mp} - T_{mp} \quad (6.1.1.3)$$

Nici pentru această variabilă aleatoare nu poate fi cunoscută repartiţia. Dar, din definiţia scorului real şi a erorii de măsură rezultă că, pentru o persoană p şi un instrument de măsură m, media erorilor de măsură este egală cu zero:

$$\mathbf{M}(E_{mp}) = 0 \quad (6.1.1.4)$$

Analizând relaţia existentă între variabilele aleatoare X_{mp}, T_{mp} şi E_{mp}:

$$X_{mp} = T_{mp} + E_{mp} \quad (6.1.1.5)$$

se poate deduce faptul că scorurile observate la un test (x_{mp}) reflectă influenţa a două feluri de factori:

- factori care contribuie la *consistenţă* (care furnizează valoarea τ_{mp}): caracteristicile stabile ale individului şi ale situaţiei în care se face măsurarea;
- factori care contribuie la *inconsistenţă* (care produc valorile e_{mp}): caracteristici ale individului sau situaţia de examinare, care afectează scorurile observate, dar care nu au nimic comun cu ceea ce se încearcă să se măsoare.

Variabilele aleatoare X_{mp}, T_{mp} şi E_{mp} au fost construite pentru un anumit instrument de măsură $m \in \mathbf{M}$ şi o persoană fixată $p \in \mathbf{P}$.

Să alegem acum un instrument de măsură $m \in \mathbf{M}$ şi să îl aplicăm, în condiţii standardizate, tuturor persoanelor din populaţia **P**. Să notăm cu X_{m*} variabila aleatoare a cărei repartiţie este dată de toate scorurile observate care s-ar putea obţine în populaţia **P**, cu probabilităţile corespunzătoare.

Vom nota cu T_{m*} variabila aleatoare care are ca valori scorurile reale ale tuturor persoanelor din populaţia **P** şi cu E_{m*} variabila aleatoare corespunzătoare erorilor de măsură care se pot obţine în populaţia **P**.

TEOREMA 6.1.1.1. (Lord, Novick, 1968). În populaţia **P** sau în orice subpopulaţie a lui **P** are loc relaţia:

$$X_{m*} = T_{m*} + E_{m*} \quad (6.1.1.6)$$

Observaţie. Dacă populaţia **P** este formată dintr-o singură persoană, $\mathbf{P}=\{p\}$, atunci egalitatea de mai sus devine

$$X_{np} = T_{np} + E_{np} \quad (6.1.1.7)$$

deci se regăseşte relaţia (6.1.1.3) pe baza căreia a fost definită variabila aleatoare E_{np}.

Din formula (6.1.1.6) se deduce că, în întreaga populaţie **P**, deosebirile observate între scorurile obţinute de subiecţi (valorile variabilei X_{m*}) provin din două surse:

- deosebirile existente între persoane în privinţa caracteristicii măsurate de test (valorile variabilei T_{m*});
- erorile de măsură (valorile variabilei E_{m*}).

Vom presupune în continuare că variabila aleatoare X_{m*} are dispersie finită:

$$\sigma^2(X_{m*}) < \infty \quad (6.1.1.8)$$

Din inegalitatea (6.1.1.8) rezultă următoarele:

$$\sigma^2(T_{m*}) < \infty \quad (6.1.1.9)$$

$$M(X_{m*}) < \infty \quad (6.1.1.10)$$

$$M(T_{m*}) < \infty \quad (6.1.1.11)$$

Pentru a elimina din discuţie cazurile banale, vom presupune că fiecare dintre variabilele X_{m*}, T_{m*} şi E_{m*} are mai multe valori (nu este constantă), ceea ce conduce la inegalităţile:

$$\sigma^2(X_{m*}) > 0 \quad (6.1.1.12)$$

$$\sigma^2(T_{m*}) > 0 \quad (6.1.1.13)$$

$$\sigma^2(E_{m*}) > 0 \quad (6.1.1.14)$$

TEOREMA 6.1.1.2 (Lord, Novick, 1968).

a. Media erorilor de măsură este egală cu zero:

$$M(E_{m*}) = 0 \quad (6.1.1.15)$$

b. Media scorurilor observate este egală cu media scorurilor reale:

$$M(X_{m*}) = M(T_{m*}) \quad (6.1.1.16)$$

c. Scorurile reale şi erorile de măsură sunt necorelate liniar:

$$\rho(E_{m*}, T_{m*}) = 0 \quad (6.1.1.17)$$

d. Pentru orice valoare particulară a scorurilor reale, τ_{m*}, media erorilor de măsură în populaţia considerată este egală cu zero:

$$M(E_{m*}/T_{m*} = \tau_{m*}) = 0 \quad (6.1.1.18)$$

(Media erorilor într-o subpopulaţie cu scor real fixat este egală cu zero.)

e. Dacă $a \in$ **M** şi $b \in$ **M** sunt două instrumente de măsură, atunci coeficientul de corelaţie liniară între scorurile reale ale unui instrument şi erorile de măsură ale celuilalt este egal cu zero:

$$\rho(E_{a*}, T_{b*}) = 0 \quad (6.1.1.19)$$

f. Se consideră două instrumente de măsură distincte, $a \in \mathbf{M}$ şi $b \in \mathbf{M}$. Dacă pentru fiecare persoană $p \in \mathbf{P}$ variabilele aleatoare X_{ap} şi X_{bp} (sau, echivalent, E_{ap} şi E_{bp}) sunt independente, atunci coeficientul de corelaţie liniară între erorile celor două instrumente de măsură este egal cu zero:

$$\rho(E_{a*}, E_{b*}) = 0 \qquad (6.1.1.20)$$

Din această teoremă se pot deduce următoarele consecinţe:

A. Pentru orice valoare posibilă a scorurilor reale, τ, media scorurilor observate este egală cu τ:

$$\mathbf{M}(X_{m*} | T_{m*} = \tau) = \tau \qquad (6.1.1.21)$$

Adică regresia scorului observat faţă de scorul real este o dreaptă care trece prin origine şi are panta egală cu unu.

B. Dispersia scorului observat este egală cu suma dispersiei scorului real şi a dispersiei erorii:

$$\sigma^2(X_{m*}) = \sigma^2(T_{m*}) + \sigma^2(E_{m*}) \qquad (6.1.1.22)$$

C. Pătratul coeficientului de corelaţie între scorurile observate şi scorurile reale este egal cu raportul dintre dispersia scorurilor reale şi dispersia scorurilor observate:

$$\rho^2(X_{m*}, T_{m*}) = \frac{\sigma^2(T_{m*})}{\sigma^2(X_{m*})} \qquad (6.1.1.23)$$

El se numeşte **coeficient de fidelitate al testului *m*.**

Coeficientul de corelaţie $\rho(X_{m*}, T_{m*})$ se numeşte **indice de fidelitate**. El arată cât de puternică este legătura liniară dintre scorurile observate şi scorurile reale pentru instrumentul de măsură $m \in \mathbf{M}$.

6.1.2. Instrumente de măsură echivalente, instrumente de măsură paralele

DEFINIŢIA 6.1.2.1 (Lord, Novick, 1968). Două instrumente de măsură distincte $a \in \mathbf{M}$ şi $b \in \mathbf{M}$ sunt **echivalente** în populaţia **P** dacă pentru orice persoană $p \in \mathbf{P}$:

- scorul real este acelaşi:

$$\tau_{ap} = \tau_{bp} \qquad (6.1.2.1)$$

şi

• erorile de măsură sunt identic distribuite, adică E_{ap} şi E_{bp} au aceeaşi repartiţie.

Dacă este îndeplinită numai condiţia 6.1.2.1, atunci instrumentele de măsură *a* şi *b* se numesc **τ-echivalente.**

DEFINIŢIA 6.1.2.2 (Lord, Novick, 1968). Instrumentele de măsură distincte $a \in M$ şi $b \in M$ sunt **esenţial τ-echivalente** dacă pentru orice persoană $p \in P$ diferenţa scorurilor reale este aceeaşi:

$$\tau_{ap} - \tau_{bp} = \text{constant} \qquad (6.1.2.2)$$

DEFINIŢIA 6.1.2.3 (Lord, Novick, 1968). Instrumentele de măsură distincte $a \in M$ şi $b \in M$ sunt **paralele** dacă pentru orice persoană $p \in P$:

• scorul real este acelaşi:

$$\tau_{ap} = \tau_{bp} \qquad (6.1.2.3)$$

şi

• erorile de măsură au aceeaşi abatere standard:

$$\sigma(E_{ap}) = \sigma(E_{bp}) \qquad (6.1.2.4)$$

Deci, două instrumente de măsură paralele măsoară exact acelaşi lucru, în aceeaşi scară ($\tau_{ap} = \tau_{bp}$) şi îl măsoară la fel de bine ($\sigma(E_{ap}) = \sigma(E_{bp})$), pentru fiecare persoană.

Relaţia de paralelism este mai slabă decât relaţia de echivalenţă, iar relaţia de τ-echivalenţă este mai slabă decât cea de paralelism:

echivalenţă => paralelism => τ-echivalenţă

Se poate demonstra următoarea condiţie necesară şi suficientă pentru ca două măsurători să fie paralele:

TEOREMA 6.1.2.1 (Lord, Novick, 1968). Două instrumente de măsură distincte, $a \in M$ şi $b \in M$, sunt paralele în populaţia P dacă şi numai dacă în orice subpopulaţie a lui P scorurile observate au aceeaşi medie şi aceeaşi dispersie:

$$M(X_{a*}) = M(X_{b*}) \qquad (6.1.2.5)$$

şi

$$\sigma^2(X_{a*}) = \sigma^2(X_{b*}) \qquad (6.1.2.6)$$

Din această teoremă se constată că două instrumente de măsură paralele într-o populaţie P sunt paralele şi în orice subpopulaţie a lui P. Dar, într-o populaţie mai mare decât P s-ar putea să nu mai fie paralele.

TEOREMA 6.1.2.2 (Lord, Novick, 1968). Fie k un număr natural mai mare decât 1, $m_1, m_2,..., m_k$ instrumente de măsură paralele din **M** şi m_o un instrument de măsură distinct, ales arbitrar. Atunci, în orice subpopulaţie a lui **P** sunt adevărate următoarele afirmaţii:

a. toate intercorelaţiile instrumentelor paralele sunt egale:

$$\rho(X_{m_i *}, X_{m_j *}) = constant \qquad (6.1.2.7)$$

$i,j=1,2,...,k, \quad i \neq j$

b. toate instrumentele de măsură paralele corelează la fel de bine cu instrumentul m_o:

$$\rho(X_{m_i *}, X_{m_0 *}) = constant \qquad (6.1.2.8)$$

$i=1,2,...,k$

Întrucât repartiţiile variabilelor aleatoare X_{mp}, pentru persoanele p din populaţia **P** şi instrumentele de măsură m din mulţimea **M**, nu sunt cunoscute, paralelismul a două instrumente nu poate fi verificat utilizând definiţia. Nici teorema care furnizează condiţiile necesare şi suficiente pentru paralelism nu poate fi folosită, deoarece nu se cunosc repartiţiile variabilelor X_{m*}. Din aceste motive, în practică, studierea paralelismului a două sau mai multe teste, m_1, $m_2,..., m_k$, se face prin verificarea, pe baza scorurilor observate ale eşantionului căruia i s-au administrat testele, a următoarelor egalităţi:

$$M(X_{m_i *}) = M(X_{m_j *}), \qquad i,j=1,2,...,k \qquad (6.1.2.9)$$

$$\sigma^2(X_{m_i *}) = \sigma^2(X_{m_j *}), \qquad i,j=1,2,...,k \qquad (6.1.2.10)$$

$$\rho(X_{m_i *}, X_{m_j *}) = constant, \qquad i,j=1,2,...,k, \quad i \neq j \qquad (6.1.2.11)$$

$$\rho(X_{m_i *}, X_{m_0 *}) = constant, \qquad i=1,2,...,k \qquad (6.1.2.12)$$

m_o fiind un test arbitrar.

Se utilizează teste statistice adecvate.

Trebuie menţionat, însă, că egalităţile de mai sus reprezintă numai condiţii necesare, nu şi suficiente pentru paralelism. Dacă ele nu sunt verificate,

atunci cu certitudine instrumentele de măsură nu sunt paralele. Dar, realizarea acestor egalităţi pentru eşantionul testat nu asigură paralelismul.

TEOREMA 6.1.2.3 (Lord, Novick, 1968). Dacă $a \in M$ şi $b \in M$ sunt două instrumente de măsură paralele şi dacă pentru orice persoană $p \in P$ scorurile observate X_{ap} şi X_{bp} sunt variabile aleatoare independente, atunci, în populaţia **P**, au loc următoarele relaţii:

$$\rho^2(X_{a*}, T_{a*}) = \rho^2(X_{b*}, T_{b*}) = \rho(X_{a*}, X_{b*}) \qquad (6.1.2.13)$$

Aceasta înseamnă că pătratul coeficientului de corelaţie liniară între scorurile reale şi scorurile observate la un test (care, reprezintă coeficientul de fidelitate al testului) este egal cu coeficientul de corelaţie liniară între scorurile observate la acest test şi la un test paralel cu el.

6.1.3. Eroarea standard de măsură

Fie ϵ un număr real pozitiv. Din inegalitatea lui Cebîşev:

$$P(|X_{mp} - \tau_{mp}| \geq \epsilon) \leq \frac{\sigma^2(E_{mp})}{\epsilon^2} \qquad (6.1.3.1)$$

se constată că dispersia erorilor de măsură pentru o persoană $p \in \mathbf{P}$, $\sigma^2(E_{mp})$, este un indicator al preciziei măsurătorii pentru respectiva persoană: atunci când $\sigma^2(E_{mp})$ este mic, există doar puţine şanse ca scorurile observate să difere de scorul real.

Trebuie remarcat că dispersia erorilor pentru o persoană este egală cu dispersia scorurilor observate ale persoanei respective:

$$\sigma^2(E_{mp}) = \sigma^2(X_{mp}) \qquad (6.1.3.2)$$

Dispersia erorilor poate să difere mult de la o persoană la alta în populaţia **P**.

Întrucât dispersia erorilor de măsură în populaţia **P**, $\sigma^2(E_{m*})$, este egală cu media dispersiilor erorilor de măsură pentru toate persoanele din **P**:

$$\sigma^2(E_{m*}) = M_p \sigma^2(E_{mp}) \qquad (6.1.3.3)$$

se consideră că disperia erorilor de măsură în populaţia **P** este un indicator al impreciziei medii a măsurătorilor în populaţia **P**.

Abaterea standard a erorilor de măsură, $\sigma(E_{m*})$, se numeşte **eroare standard de măsură** în **P**. În continuare o vom nota cu σ_E:

$$\sigma_E = \sigma(E_{m^*}) \qquad (6.1.3.4)$$

6.2. TEORIA RĂSPUNSULUI LA ITEMI

Teoria răspunsului la itemi nu este cu adevărat o teorie; ea reprezintă o colecţie de modele şi metode statistice cu ajutorul cărora se dă un sens datelor provenite din măsurători psihologice (Steinberg, Thissen, 1996).

6.2.1. Variabile latente, variabile observate

Teoria răspunsului la itemi are la bază următoarele postulate (Hambleton, Swaminathan, Rogers, 1991; Steinberg, Thissen, 1996):

a. Comportarea unui subiect la un item al unui test poate fi explicată (sau prezisă) cu ajutorul unui set de *factori* numiţi **trăsături, trăsături latente** sau **abilităţi**. Orice construct inobservabil, presupus continuu, în privinţa căruia o teorie psihologică afirmă că persoanele se deosebesc între ele poate fi privit ca factor. Fiecărui factor i se asociază o variabilă cu valori numerice, cuprinse între -∞ şi +∞, continuă, denumită **variabilă latentă.**

b. Se infereaza existenţa unui factor numai după ce se observă că răspunsurile la itemii prin care se doreşte măsurarea constructului covariază (variază la fel).

c. Relaţia dintre performanţa la item a subiecţilor şi fiecare dintre trăsăturile care are legătură cu performanţa poate fi descrisă prin câte o funcţie crescătoare denumită **funcţie caracteristică a itemului** sau **curbă caracteristică a itemului**. Această funcţie arată cum depinde probabilitatea de a răspunde corect (sau afirmativ) la item de nivelul trăsăturii.

O trăsătură nu poate fi nici observată, nici măsurată direct. Măsura ei se obţine indirect, prin intermediul uneia sau al mai multor **variabile observate** (de exemplu, răspunsurile la itemii unui test sau chestionar), despre care se presupune că reprezintă adecvat variabila latentă.

Operarea cu variabile latente impune rezolvarea a trei probleme (Andrich, 1988):

A. ***Definirea variabilei latente în termeni operaţionali***, care să permită măsurarea ei, acţiune denumită *construirea variabilei.*

Adesea, pentru construirea unei variabile se identifică acele diferenţe individuale care fac posibilă reprezentarea fiecărui individ printr-un punct pe

axa reală. Despre o asemenea variabilă se spune că este **unidimensională**. Dacă, însă, o reprezentare acceptabilă a caracteristicilor considerate necesită utilizarea unui spaţiu cu mai multe dimensiuni, se spune că variabila este **multidimensională**.

Orice performanţă umană, acţiune sau credinţă este complexă şi implică o multitudine de aptitudini, interese etc. Reducerea ei la o singură dimensiune poate fi realizată în mai multe moduri. Din acest motiv, o variabilă unidimensională reflectă concepţia celui care o construieşte. Pentru o aceeaşi caracteristică psihică, doi autori diferiţi pot propune variabile diferite.

B. ***Definirea situaţiilor controlabile*** care permit obţinerea de informaţii valide şi repetate asupra variabilei cercetate, operaţie denumită *crearea cadrului de observaţii*. Asemenea situaţii sunt reprezentate de itemii de test.

Rezultatul observaţiilor este exprimat numeric, printr-un scor.

C. ***Alegerea modelului matematic*** pentru obţinerea valorii variabilei latente pe baza valorilor variabilelor observate.

În majoritatea cazurilor, itemii conţin afirmaţii sau întrebări la care există un număr redus de variante de răspuns (de exemplu, "da"/"nu", "rareori"/"uneori"/"adesea"). Deci fiecare item furnizează un scor dintr-o mulţime discretă, cu puţine elemente. Scorul total al testului, calculat pe baza scorurilor itemilor, ia de asemenea valori dintr-o mulţime discretă.

Variabila latentă, prin definiţie, este continuă şi ia valori pe întreaga axă reală.

Deci, modelul matematic ales trebuie să transforme o mulţime discretă (mulţimea scorurilor observate la un item sau la test) în mulţimea numerelor reale.

Variabilele latente apar în două tipuri de modele utilizate pentru reprezentarea legăturilor dintre variabilele psihologice (Ozer, Reise, 1994):

- În modelele care măsoară cauzele se presupune că variabila latentă este cauza variaţiilor variabilelor observate. Acest model stă la baza standardelor actuale, care cer ca o măsură bună a unui construct să aibă o consistenţă internă ridicată.
- În modelele care măsoară efectele, variabila latentă este cauzată de variabilele observate. Uneori, variabila efect este privită ca o combinaţie liniară a mai multor variabile observate (aşa se întâmplă, de exemplu, pentru leadership). Alteori, ca, de exemplu, pentru extraversiune, relaţia dintre cauze şi efect nu poate fi exprimată printr-o ecuaţie de regresie liniară multiplă.

Dacă se utilizează mai multe variabile observate (instrumente de măsură) pentru cercetarea aceleiaşi variabile latente, se presupune că în spatele

covariaţiilor existente între variabilele observate stau relaţiile dintre aceste variabile şi variabila latentă pe care o măsoară. Deci se consideră că variabila latentă explică relaţiile dintre variabilele observate.

Această idee a stat la baza utilizării analizei factoriale ca model matematic ce permite determinarea valorilor unuia sau al mai multor factori (variabile latente) pornind de la mai multe variabile observate (teste).

Prin analiză factorială, pentru fiecare persoană se înlocuiesc scorurile realizate la teste (valorile variabilelor observate) cu scorurile factoriale (valorile variabilelor latente măsurate de teste). Aceste scoruri factoriale sunt utilizate pentru caracterizarea persoanei sau pentru efectuarea de predicţii, prin introducerea lor în calculul valorilor unor funcţii de regresie.

Administrând unei populaţii un set de itemi dihotomici despre care se presupune că măsoară aceeaşi variabilă latentă unidimensională se pot face următoarele două observaţii importante:

- persoane diferite răspund diferit la itemi, ceea ce se poate explica prin deosebirea dintre indivizi în privinţa variabilei latente măsurate;
- o aceeaşi persoană dă răspunsuri diferite la itemi diferiţi, deşi aceştia măsoară aceeaşi variabilă latentă; explicaţia posibilă ar fi că itemii măsoară în mod diferit variabila latentă. L.L. Thurstone a denumit această caracteristică a itemilor "valoare afectivă" (Andrich, 1988).

Pentru modelul matematic care leagă valorile variabilei observate corespunzătoare unui item, de valorile variabilei latente a fost propusă ecuaţia următoare (Andrich, 1988):

$$P(X=x|p)=f(Y(p), B, x) \qquad (6.2.1.1)$$

unde:

X este variabila aleatoare corespunzătoare itemului considerat;
$x \in \{0,1\}$ este un scor posibil la item;
p este o persoană care a răspuns la item;
$Y(p)$ reprezintă valoarea variabilei latente pentru persoana p;
B este valoarea afectivă a itemului;
f este o funcţie, a cărei expresie analitică trebuie precizată.

Deci, probabilitatea ca o persoană p să dea un anumit răspuns la item ($P(X=x|p)$) depinde atât de persoană ($Y(p)$), cât şi de item (B).

Făcând următoarele ipoteze:

$$Y(p)>0 \text{ pentru orice persoană } p \qquad (6.2.1.2)$$
$$B>0 \qquad (6.2.1.3)$$

Rasch a propus pentru funcţia f expresia (Andrich, 1988):

$$f(Y,B,x) = \left\{\frac{\frac{Y}{B}-1}{1+\frac{Y}{B}}\right\}x + \frac{1}{1+\frac{Y}{B}} \qquad (6.2.1.4)$$

Se observă că, atunci când raportul Y/B este mare (ceea ce se întâmplă când nivelul subiectului în privinţa variabilei latente măsurate de item este ridicat sau valoarea afectivă a itemului este scăzută) probabilitatea de obţinere la item a scorului 1 este mare.

Din relaţia

$$\frac{Y(p)}{B} = \frac{P(X=1|p)}{P(X=0|p)} \qquad (6.2.1.5)$$

ţinând seama că pentru un item fixat valoarea B este constantă, se deduce că valoarea variabilei latente pentru o persoană, $Y(p)$, este proporţională cu raportul dintre probabilitatea ca persoana respectivă să obţină la item scorul 1 şi probabilitatea de a obţine scorul 0. Cu cât variabila latentă are o valoare mai mare, cu atât mai mare este probabilitatea persoanei de a obţine scorul 1, comparativ cu probabilitatea de a obţine scorul 0.

Să presupunem că doi itemi, i şi j, având valorile afective B_i şi, respectiv, B_j, sunt independenţi probabilistic între ei, pentru orice persoană p. Aceasta înseamnă că oricare ar fi două scoruri posibile la itemi, $x,y \in \{0,1\}$ are loc egalitatea

$$\mathrm{P}((X_i=x)\cap(X_j=y)\mid p) = \mathrm{P}((X_i=x)\mid p)\,\mathrm{P}((X_j=y)\mid p) \qquad (6.2.1.6)$$

În acest caz se poate deduce relaţia

$$\frac{P((X_i=1) \bigcap (X_j=0)|p)}{P((X_i=0) \bigcap (X_j=1)|p} = \frac{B_j}{B_i} \qquad (6.2.1.7)$$

Deci, pentru orice persoană, indiferent de valoarea variabilei latente a acesteia, raportul dintre probabilitatea de a obţine scorul 1 numai la itemul i şi probabilitatea de a obţine scorul 1 numai la itemul j este aceeaşi, dependentă doar de caracteristicile celor doi itemi.

Introducând notaţiile:

$y = \ln(Y)$ (6.2.1.8)

$$b = \ln(B) \tag{6.2.1.9}$$

şi utilizând relaţia

$$\ln(\frac{Y}{B}) = y - b \tag{6.2.1.10}$$

formula de definiţie a funcţiei *f*, (6.2.1.4) se transcrie astfel:

$$P(X=x|p) = \frac{1 + x[e^{y(p)-b} - 1]}{1 + e^{y(p)-b}} \tag{6.2.1.11}$$

Valoarea *y*(*p*) reflectă nivelul variabilei latente pentru persoana *p* (*y* fiind imaginea valorii variabilei latente, *Y*, printr-o funcţie crescătoare), iar valoarea *b* caracterizează itemul. La un nivel *y* fixat, probabilitatea de a obţine scorul 1 este cu atât mai mare, cu cât parametrul *b* este mai mic.

Scopul teoriei răspunsului la itemi este de a elabora metode de estimare a valorii variabilei latente la persoanele testate şi a caracteristicilor itemilor testului, pe baza răspunsurilor date de subiecţi la itemi.

6.2.2. Funcţia caracteristică a unui item

În cele ce urmează ne vom referi doar la itemi dihotomici şi vom utiliza expresia "itemul a fost rezolvat corect (respectiv greşit)" cu înţelesul "s-a obţinut la item scorul 1 (respectiv 0)".

Vom presupune că toţi itemii testului măsoară aceeaşi variabilă latentă unidimensională.

Teoria răspunsului la itemi impune în acest caz ipoteza unidimensionalităţii spaţiului variabilelor latente, ceea ce înseamnă că nu există o altă variabilă latentă, pe lângă cea specificată, care să influenţeze scorurile itemilor.

Această ipoteză asigură îndeplinirea "**condiţiei de independenţă locală a itemilor**", care cere ca modul în care o persoană răspunde la un item să nu fie influenţat de felul în care a răspuns la ceilalţi itemi (De Gruijter, Van der Kamp, 1984). Trebuie menţionat că presupunerea de independenţă locală a itemilor nu implică faptul că itemii sunt necorelaţi în întregul lot de subiecţi. Dacă persoanele examinate diferă în privinţa variabilei latente măsurate de itemi, corelaţiile dintre itemi vor fi diferite de zero.

Teoria răspunsului la itemi construieşte, pentru fiecare item al testului, **funcţia caracteristică a itemului** (Lord, 1980; De Gruijter, Van der Kamp, 1984). Această funcţie leagă valorile variabilei latente măsurate de item cu scorurile obţinute la item. Ea este definită astfel:

$$P : \mathrm{R} \to [0,1]$$

$P(y)$ = probabilitatea ca o persoană având nivelul variabilei latente măsurate de item egal cu y să răspundă corect la item,
pentru orice valoare posibilă, $y \in \mathrm{R}$.

Cu ajutorul acestei funcţii pot fi apreciate calităţile itemului (dificultatea şi puterea de discriminare). Graficul său realizează o vizualizare a comportării itemului ca instrument de măsură.

Teoria răspunsului la itemi face ipoteza că funcţia P este crescătoare (Lord, 1980). Într-adevăr, dacă itemul este un instrument de măsură bun, este logic ca o persoană care are un nivel mai înalt al variabilei latente să aibă şanse mai mari de a rezolva corect itemul decât un subiect care are un nivel mai scăzut.

Se mai presupune că toţi itemii testului au funcţiile caracteristice din aceeaşi clasă de funcţii.

Dintre clasele de funcţii propuse, un model considerat ca fiind foarte apropiat de ceea ce se întâlneşte obişnuit în practică este cel logistic (Lord, 1980), care utilizează funcţii de forma următoare :

$$P(y) = c + \frac{1-c}{1+e^{-1{,}7a(y-b)}} \qquad (6.2.2.1)$$

a, b şi c sunt parametri care caracterizează itemul, iar e este constanta 2.71828. Graficul funcţiei P este prezentat în figura 6.2.2.1.

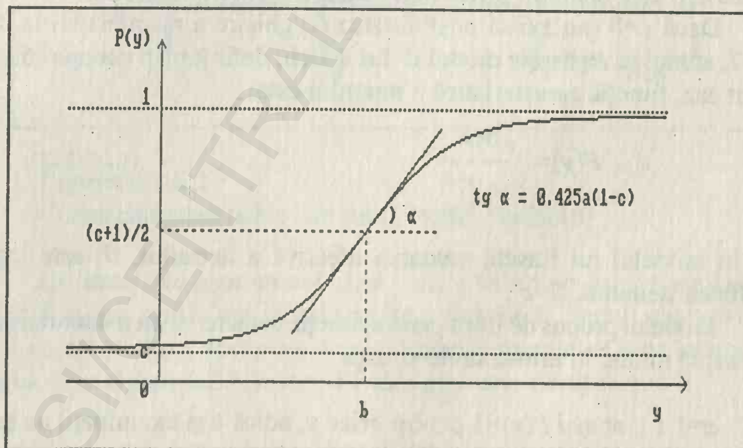


Figura 6.2.2.1
Funcţia caracteristică a itemului (model Lord)

Parametrul c reprezintă probabilitatea ca o persoană complet lipsită de caracteristica măsurată de item (y=-∞) să răspundă corect. Se numeşte **parametru de ghicire** sau **nivelul scorului pseudoşansei**. Dacă la un item nu se poate răspunde corect prin ghicire, atunci c=0.

Parametrul b, numit **dificultatea itemului**, determină poziţia curbei: cu cât itemul este mai dificil (b este mai mare), cu atât curba are porţiunea cu creştere mai rapidă deplasată mai mult spre dreapta. Funcţia P are un punct de inflexiune pentru y=b.

Dacă nu se poate ghici răspunsul (c=0), atunci b reprezintă nivelul variabilei latente, y, pentru care şansa de a răspunde corect este de 50%. Dacă răspunsul corect poate fi ghicit ($c \neq 0$), atunci b reprezintă nivelul caracteristicii y la care probabilitatea răspunsului corect se află la jumătatea distanţei dintre c şi 1:

$$P(b)=\frac{c+1}{2}$$

Parametrul a este proporţional cu panta funcţiei caracteristice a itemului în punctul de inflexiune. Mai precis, coeficientul unghiular al tangentei la curbă în punctul y=b este egal cu $0.425a(1-c)$. Aceasta înseamnă că viteza cu care variază probabilitatea de a răspunde corect atunci când nivelul variabilei latente este egal cu b, este cu atât mai mare, cu cât a este mai mare şi c este mai mic. Parametrul a este denumit **puterea de discriminare** a itemului.

Dacă c=0 (nu există posibilitatea de ghicire a răspunsului la item) şi a=1/1.7, atunci se regăseşte modelul lui Rasch, definit prin ecuaţia (6.2.1.11). În acest caz, funcţia caracteristică a itemului este:

$$P(y)=\frac{e^{(y-b)}}{1+e^{(y-b)}} \qquad (6.2.2.2)$$

Deci, în modelul lui Rasch, valoarea afectivă a itemului, B, este legată de dificultatea itemului: $B=e^b$.

Modelul propus de Lord pentru funcţia caracteristică a itemului conduce la o dreaptă numai în următoarele situaţii :

c=1 : atunci $P(y)$=1 pentru orice y, adică toţi examinaţii au şansa de 100% să răspundă corect;

a=0 : atunci $P(y)=(c+1)/2$ pentru orice y.

In modelul propus de Lord funcţia caracteristică a itemului ia valori în intervalul (c,1), având asimptotele orizontale y=c şi y=1. Deci niciodată

probabilitatea de a răspunde corect nu va putea fi egală cu c sau cu 1. Aceste valori sunt atinse doar pentru $y=-\infty$ şi, respectiv, $y=+\infty$.

Guilford (1965) imaginează un alt model pentru funcţia caracteristică a itemului, fără a propune, însă, forma analitică a funcţiei P. El consideră că funcţia P este definită şi nedescrescătoare pe un anumit interval $[y_{min}, y_{max}]$, cu proprietatea că există o valoare $y_0 \geq y_{min}$ sub care nici o persoană nu rezolvă corect itemul:

$$P(y)=0 \text{ pentru } y \leq y_0$$

şi există o valoare y_1, $y_0 \leq y_1 \leq y_{max}$, dincolo de care toată lumea răspunde corect la item:

$$P(y)=1 \text{ pentru } y \geq y_1.$$

Intervalul $[y_0, y_1]$ se numeşte "zonă de tranziţie".

Modelul propus de Guilford pentru curba caracteristică a itemului este reprezentat grafic în figura 6.2.2.2.

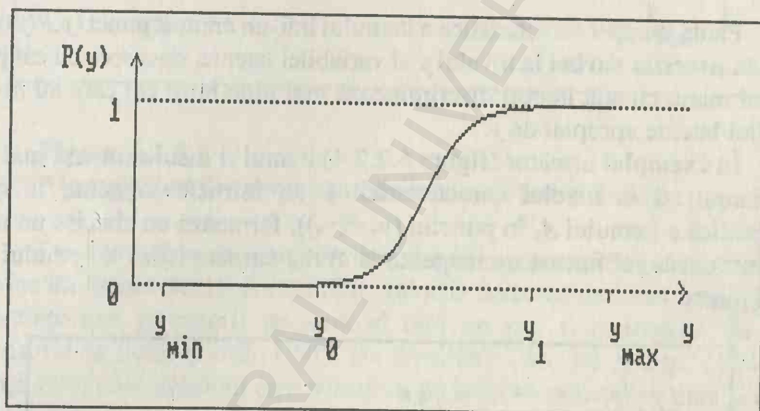


Figura 6.2.2.2
Curba caracteristică a itemului (model Guilford)

Un item este uşor atunci când pentru valori ale lui y apropiate de y_{min} funcţia P creşte rapid, apropiindu-se de valoarea 1. Itemul este dificil dacă pentru valori ale lui y apropiate de y_{max} valorile funcţiei P sunt îndepărtate de valoarea 1. În figura 6.2.2.3 itemul 1 este mai uşor decât itemul 2.

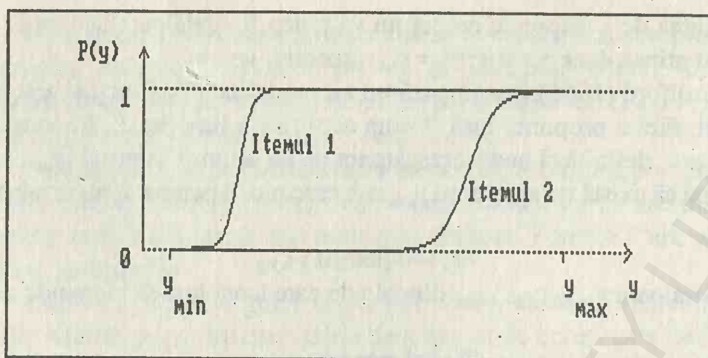


Figura 6.2.2.3
Curbele caracteristice pentru doi itemi de dificultăţi diferite

Panta funcţiei caracteristice a itemului într-un anumit punct $(y,P(y))$ este denumită **precizia curbei** la nivelul y al variabilei latente, deoarece cu cât panta este mai mare, cu atât itemul discriminează mai bine între cei care au nivelul variabilei latente apropiat de y.

În exemplul următor (figura 6.2.2.4) itemul A discriminează mai bine decît itemul B la nivelul caracteristicii $y=y_*$, întrucât tangenta la curba caracteristică a itemului A, în punctul $(y_*,P(y_*))$, formează cu abscisa un unghi mai mare decât cel format de tangenta la curba caracteristică a itemului B în acelaşi punct.

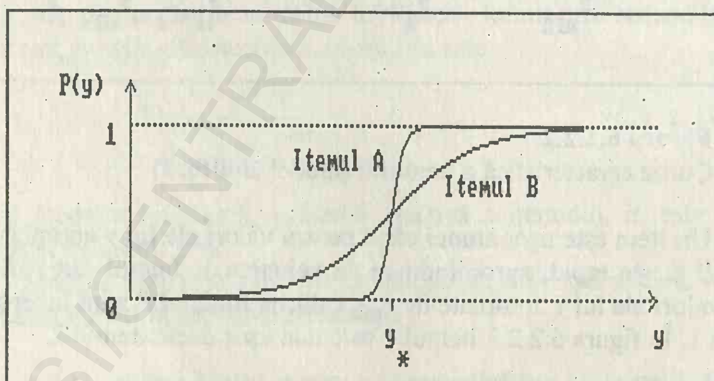


Figura 6.2.2.4
Curbele caracteristice pentru doi itemi cu precizii diferite la nivelul y_*

Guilford (1965) propune ca la construcţia unui test să se aleagă itemi având o precizie mare în zona de tranziţie şi ale căror zone de tranziţie să ocupe poziţii diferite în intervalul $[y_{min}, y_{max}]$ (figura 6.2.2.5). În felul acesta s-ar obţine un test cu o bună capacitate de discriminare între persoane.

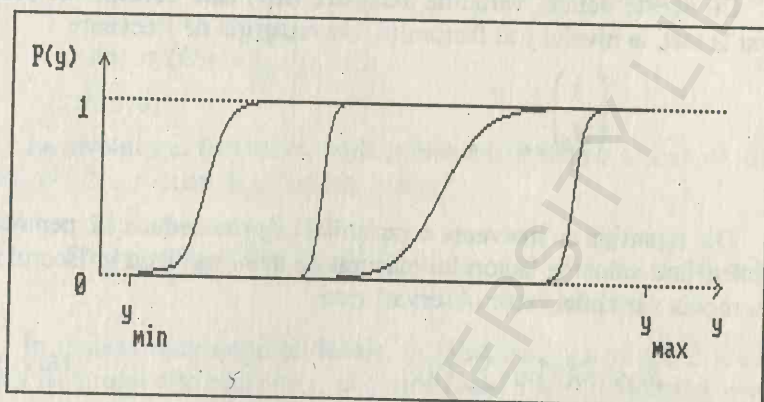


Figura 6.2.2.5
Alegerea itemilor pentru construirea unui test psihologic

La determinarea parametrilor funcţiei caracteristice a unui item intervine un aspect deosebit de dificil: valorile caracteristicii măsurate de item (cunoştinţe sau priceperi) nu sunt şi nici nu pot fi cunoscute. În teoria răspunsului la itemi (Lord, 1980; De Gruijter, Van der Kamp, 1984) sunt descrişi, însă, unii algoritmi care permit ca pe baza răspunsurilor date la un test de către un lot de persoane să se estimeze atât parametrii itemilor testului (coeficienţii care apar în expresia analitică a funcţiilor caracteristice), cât şi valorile caracteristicii măsurate de itemi, la persoanele examinate.

6.2.3. Utilizarea rezultatelor la un test

Să considerăm un test format din itemii $i_1, i_2, ..., i_n$, cu funcţiile caracteristice la item, P_j, $j=1,2,...,n$, cunoscute. Scorul testului se calculează prin însumarea scorurilor itemilor, deci ia valori din mulţimea $\{0,1,...,n\}$. Probabilitatea ca o persoană care are valoarea factorului măsurat de itemi egală cu y să obţină la test scorul $k \in \{0,1,...,n\}$ este egală cu coeficientul termenului x^k, A_k, din polinomul

$$R(x;y)=\prod_{j=1}^{n}\{P_j(y)x+[1-P_j(y)]\} = \sum_{j=0}^{n} A_j x^j \qquad (6.2.3.1)$$

Cu aceste notaţii, variabila aleatoare $(X|y)$ care corespunde scorului observat la test, la nivelul y al factorului, are repartiţia de frecvenţe

$$\begin{pmatrix} k \\ A_k \end{pmatrix}_{k=0,1,\ldots,n} \qquad (6.2.3.2)$$

Din repartiţia de frecvenţe a variabilei $(X|y)$ se deduce că, pentru orice persoană având valoarea factorului măsurat de itemi egală cu y, "scorul real", egal cu media variabilei "scor observat" este

$$S(y)=M(X|y)=\sum_{j=1}^{n} P_j(y) \qquad (6.2.3.3)$$

Înseamnă că toate persoanele care au aceeaşi valoare a factorului au acelaşi scor real. Adică, scorul real şi valoarea factorului sunt acelaşi lucru, dar măsurat în scale diferite. Deosebirea esenţială între ele constă în faptul că scorul real depinde de caracteristicile itemilor testului.

Funcţiile P_j, $j=1,2,\ldots,n$ fiind crescătoare, rezultă că şi funcţia scor real S este crescătoare.

Diferenţa dintre scorul observat şi scorul real al unei persoane se numeşte **eroare de măsură**. Ea poate fi privită ca o valoare a variabilei aleatoare

$$(E \mid y)=(X \mid y)-\mathrm{M}(X \mid y) \qquad (6.2.3.4)$$

y fiind valoarea factorului pentru persoana respectivă. Abaterea standard a variabilei $(E \mid y)$ se numeşte **eroare standard de măsură** şi este un indicator al preciziei testului, arătând cât de mult se abat scorurile observate de scorul real.

Spre deosebire de teoria clasică a testelor, unde eroarea standard de măsură este constantă pentru întreaga populaţie, în teoria răspunsului la itemi eroarea standard de măsură variază în funcţie de nivelul factorului măsurat de test, y.

Din legea de repartiţie a variabilei $(X|y)$ se deduce

$$\sigma^2(E|y)=\sum_{j=1}^{n} P_j(y)[1-P_j(y)] \qquad (6.2.3.5)$$

Dacă P_j, $j=1,2,\ldots,n$, sunt funcţii logistice, atunci:

$$\lim_{y \to \infty} \sigma^2(E|y)=0 \qquad (6.2.3.6)$$

(la niveluri foarte înalte ale factorului testul are o precizie foarte bună) şi

$$\lim_{y \to -\infty} \sigma^2(E|y)=\sum_{i=1}^{n} c_j(1-c_j) \qquad (6.2.3.7)$$

La nivelul y al factorului, probabilitatea de a obţine scorul $u_j \in \{0,1\}$ la itemul i_j, $j=1,2,...,n$ poate fi scrisă sub forma

$$P(u_j|y)=[P_j(y)]^{u_j}[1-P_j(y)]^{1-u_j} \qquad (6.2.3.8)$$

În ipoteza independenţei locale, probabilitatea ca un subiect, a cărui valoare a factorului este egală cu y, să obţină la cei n itemi ai testului scorurile $u_1, u_2,...,u_n$ (u_j este scorul la itemul i_j, $j=1,2,...,n$) este egală cu

$$P(u_1,u_2,...,u_n|y)=\prod_{i=1}^{n} [P_j(y)]^{u_j}[1-P_j(y)]^{1-u_j} \qquad (6.2.3.9)$$

Una dintre problemele la care trebuie să răspundă teoria răspunsului la itemi este aceea de a identifica nivelul factorului măsurat de test, pentru un subiect care la itemii i_j, $j=1,2,...,n$, cu funcţiile caracteristice $_jP$, $j=1,2,...,n$, cunoscute, a obţinut scorurile u_j, $j=1,2,...,n$. O modalitate de rezolvare constă în determinarea valorii y care maximizează probabilitatea $P(u_1,u_2,....,u_n|y)$. În teoria răspunsului la itemi se rezolvă ecuaţia

$$\frac{\partial \ln\ P(u_1,u_2,...,u_n|y)}{\partial y}=0 \qquad (6.2.3.10)$$

Aceasta se transcrie sub forma:

$$\sum_{j=1}^{n} \frac{P_j(y)}{P_j(y)[1-P_j(y)]}[u_j-P_j(y)] =0 \qquad (6.2.3.11)$$

Se impune cercetarea existenţei şi a unicităţii soluţiei ecuaţiei (6.2.3.11) precum şi a modalităţii de rezolvare. De exemplu, dacă toate funcţiile P_j,

j=1,2,...,n, sunt funcţii logistice cu aceiaşi parametri a, b şi c, P_j=P, j=1,2,...,n, atunci ecuaţia (6.2.3.11) devine:

$$P(y)=\frac{1}{n}\sum_{j=1}^{n} u_j \qquad (6.2.3.12)$$

Dacă c>0, iar u_1=u_2=...=u_n=0 ecuaţia (6.2.3.12) nu are soluţie.

Problema cea mai dificilă a teoriei răspunsului la itemi o constituie determinarea funcţiilor răspunsului la item, P_j, respectiv a parametrilor a_j, b_j şi c_j, j=1,2,...,n, întrucât la prima administrare a testului nu se cunosc nici nivelurile factorului pentru persoanele examinate şi nici aceşti parametri.

Se notează cu $u_{1,s}$, $u_{2,s}$,...,$u_{n,s}$ cotele celor n itemi pentru fiecare subiect s=1,2,...,S care a rezolvat testul şi se construieşte funcţia

$$L(a,b,c,y)=\prod_{s=1}^{S}\prod_{j=1}^{n}[P_j(y_s)]^{u_{j,s}}[1-P_j(y_s)]^{1-u_{j,s}} \qquad (6.2.3.13)$$

unde a=(a_1,a_2,...,a_n), b=(b_1,b_2,...,b_n), c=(c_1,c_2,...,c_n), y=(y_1,y_2,...,y_S).

Pentru determinarea vectorilor necunoscuţi a, b, c şi y se rezolvă sistemul:

$$\begin{cases} \dfrac{\partial L(a,b,c,y)}{\partial a_j}=0, & j=1,2,...n \\ \dfrac{\partial L(a,b,c,y)}{\partial b_j}=0, & j=1,2,...,n \\ \dfrac{\partial L(a,b,c,y)}{\partial c_j}=0, & j=1,2,...,n \\ \dfrac{\partial L(a,b,c,y)}{\partial y_s}=0, & s=1,2,..,.,S \end{cases}$$

(6.2.3.14)

Şi acest sistem trebuie studiat în privinţa existenţei şi a unicităţii soluţiei.

Teoria răspunsului la itemi este utilizată mai ales în testarea adaptativă.

Testele adaptative utilizează, pentru selectarea itemului care se administrează în fiecare moment, **funcţia de informaţie a itemului**. Aceasta este definită prin formula (Lord, 1980):

$$I(y) = \frac{\{P'(y)\}^2}{P(y)\{1-P(y)\}} \qquad (6.2.3.15)$$

unde P este funcţia caracteristică a itemului, iar P' reprezintă derivata funcţiei P.

Valoarea $I(y)$ este invers proporţională cu pătratul erorii standard cu care y poate fi măsurat de item (Weiss, Vale, 1987). Deci, un item aduce informaţia maximă la acel nivel y pentru care $I(y)$ este maxim.

În modelul lui Lord, la un nivel y fixat, valoarea funcţiei de informaţie a itemului, $I(y)$ este direct proporţională cu pătratul puterii de discriminare a itemului, a. Deci informaţia adusă de item este cu atât mai multă, cu cât itemul discriminează mai bine.

Pornind de la o mulţime de itemi ai căror parametri sunt cunoscuţi, se poate construi un test adaptativ după următorul algoritm (Stocking, 1987; Weiss, Vale, 1987; Schoonman, 1989):

(1) Se alege primul item, aleator, dintr-un grup de itemi care au valoarea maximă a funcţiei de informaţie la un nivel mediu al variabilei latente măsurate.

(2) Dacă acest item a fost rezolvat corect (respectiv, greşit), se alege al doilea item, aleator, dintr-o mulţime de itemi ale căror funcţii de informaţie au valoarea maximă pentru o valoare ridicată (respectiv, scăzută) a variabilei latente.

(3) Se estimează valoarea variabilei latente, y_*.

Dacă este satisfăcut criteriul de stop, se opreşte testarea, în caz contrar se alege un item care are valoarea maximă a funcţiei de informaţie în y_* şi se reia de la pasul 3.

6.3. TEORIA STĂRILOR ŞI A TRĂSĂTURILOR LATENTE

6.3.1. Ideea de bază a teoriei stărilor şi a trăsăturilor latente

Descrierea şi explicarea stabilităţii şi a schimbării în comportamentul uman sunt sarcinile de bază ale teoriei psihologice şi ale evaluării psihologice.

Orice atribut psihologic poate fi analizat cel puţin din două puncte de vedere, şi anume:

- al diferenţelor *inter*individuale;
- al diferenţelor *intra*individuale (sau al schimbărilor care au loc la fiecare individ).

Disciplinele tradiţionale ale psihologiei sunt interesate în principal de câte un singur tip de diferenţe.

Psihologia diferenţială îşi concentrează atenţia asupra deosebirilor dintre persoane, adică asupra diferenţelor *inter*individuale.

Psihologia generală (de exemplu învăţarea, percepţia), psihologia socială şi psihologia dezvoltării se ocupă de descrierea şi explicarea deosebirilor de comportament ale aceleiaşi persoane de la o situaţie la alta şi de la un moment la altul, adică a deosebirilor *intra*individuale.

Fiecare din aceste perspective consideră că celelalte surse de varianţă sunt mai puţin importante sau mai puţin interesante pentru explicarea comportamentului.

Teoria stărilor şi a trăsăturilor recunoaşte pentru atributele psihologice existenţa atât a deosebirilor dintre indivizi, stabile în timp, cât şi a schimbărilor intraindividuale. Ea consideră că fiecare atribut psihologic observat este afectat într-un anumit grad de:

- caracteristici ale individului;
- caracteristici ale situaţiei şi/sau ale influenţelor care interacţionează;
- eroarea de măsură.

Prin urmare, rezultatul măsurării unei variabile observate se descompune în:

(1) o componentă care nu depinde de situaţie şi/sau de efectele interacţiunilor;

(2) o componentă care depinde de situaţie şi/sau de interacţiuni;

(3) o eroare de măsură.

Prima componentă este denumită **trăsătură**; suma primelor două componente este denumită **stare**.

Existenţa componentei (2) are implicaţii importante asupra evaluării psihologice. Construirea şi utilizarea instrumentelor de măsură impune să se cunoască nu numai fidelitatea acestora ci şi cât de mult ele măsoară caracteristici stabile ale persoanelor şi cât de mult ele sunt afectate de situaţia în care se face măsurarea. Când se urmăreşte măsurarea *trăsăturilor*, rezultatele măsurărilor trebuie să fie afectate cât mai puţin de efectele specifice ocaziei de măsurare. Când se evaluează *starea*, instrumentul utilizat trebuie să fie sensibil la influenţele situaţiei.

6.3.2. Scopul teoriei stărilor şi a trăsăturilor latente

Teoria stărilor şi a trăsăturilor încearcă să dea răspuns la următoarele întrebări:

• Cum putem afla dacă în răspunsurile date de un individ la un chestionar care măsoară o trăsătură de personalitate intervin sau nu efectele situaţiei de măsurare ?

• Dacă există efecte ale situaţiei, ce relaţie există între scorul obţinut la scală şi trăsătura măsurată ?

• Ce relaţie există între stări şi trăsături ?

6.3.3. Variabile utilizate de teoria stărilor şi a trăsăturilor latente

Vom prezenta modelul matematic propus de teoria stărilor şi al trăsăturilor prin comparaţie cu cel utilizat de teoria clasică a testelor. Pentru aceasta vom relua ideile de bază ale teoriei clasice a testelor în formularea dată de D. W. Zimmerman (Steyer, 1989).

În **teoria clasică a testelor** se lucrează cu experimente aleatoare de următorul tip:

Se presupune că o mulţime de persoane **P** a fost examinată cu un chestionar compus din *m* itemi. Notând cu S_i mulţimea răspunsurilor posibile la itemul *i*, *i*=1,2,...,*m*, produsul cartezian

$$M=S_1xS_2x...xS_m$$

reprezintă mulţimea răspunsurilor posibile la chestionar.

De exemplu, dacă un chestionar este compus din 3 itemi cu răspunsurile posibile "da" şi "nu", atunci $S_1=S_2=S_3=\{"da","nu"\}$.

Rezultatele posibile ale experimentului aleator considerat în teoria clasică a testelor formează produsul cartezian

$$Z = PxM = Px(S_1xS_2x...xS_m)$$

Fiecare element al său $(p,s_1,s_2,...,s_m)\in Z$ este constituit dintr-o persoană, $p\in P$, şi *m* rezultate la itemi $s_i\in S_i$, *i*=1,2,...,*m*.

Se notează

$p_0 : Z \to P$ operatorul care ia ca valori persoanele din P;

$Y_i : Z \to R$ variabila care ia ca valori scorurile la itemul *i*, *i*=1,2,...,*m*.

Fiecare variabilă Y_i, *i*=1,2,...,*m*, este o variabilă aleatoare. În mulţimea de persoane **P**, ea ia fiecare valoare cu o anumită probabilitate.

Dacă în exemplul anterior fiecare item este cotat cu 1 pentru răspunsul "da" şi cu 0 pentru răspunsul "nu", atunci fiecare din variabilele Y_1, Y_2, Y_3 ia valorile 0 şi 1 cu anumite probabilităţi (care, de obicei, sunt diferite de la o variabilă la alta).

Pentru fiecare persoană $p\in P$ şi pentru fiecare item *i*=1,2,...,*m*, se definesc variabilele:

• **scor observat**, care se obţine din variabila Y_i punând condiţia ca din mulţimea **P** să fie aleasă persoana p: $(Y_i | p_0=p)$;

• **scor real**, T_i, ca fiind media variabilei aleatoare Y_i condiţionată de alegerea persoanei p din mulţimea **P**:

$$T_i = \mathbf{M}(Y_i | p_0=p) \qquad (6.3.3.1)$$

Adică T_i este media tuturor valorilor de scor care ar fi putut fi obţinute de persoana p la itemul i, dacă s-ar fi făcut toate măsurătorile posibile.

Variabilele **eroare**, E_i, pentru persoana p şi fiecare item i, i=1,2,...,m, se definesc ca fiind diferenţele dintre scorul observat şi scorul real la item:

$$E_i = (Y_i | p_0=p) - T_i \qquad (6.3.3.2)$$

În **teoria stărilor şi a trăsăturilor** se presupune că există n ocazii de măsurare, $n \geq 1$ (de obicei $n \geq 2$).

În locul mulţimii de persoane **P** se consideră:

• o mulţime de persoane **P** şi

• n mulţimi de situaţii posibile U_k, k=1,2,...,n, unde U_k reprezintă mulţimea tuturor situaţiilor care ar putea să apară în cea de a k-a ocazie de măsurare.

De exemplu, la prima administrare a unui chestionar persoanele ar putea fi "obosite", "plictisite" sau "nervoase".

Mulţimile U_k, k=1,2,...,n nu sunt cunoscute, dar se presupune că ele există.

Cu alte cuvinte, spre deosebire de teoria clasică a testelor, unde sunt măsurate *persoanele*, în teoria stărilor şi a trăsăturilor sunt măsurate *persoanele în situaţii*.

Pentru fiecare ocazie de măsurare, k=1,2,...,n, se notează cu $\mathbf{M}_k$ mulţimea rezultatelor posibile. De exemplu, dacă se administrează de n=3 ori un chestionar format din patru itemi ale căror răspunsuri posibile sunt (în paranteză, după fiecare răspuns, este notat scorul acordat):

• pentru itemul 1 : "da" (1), "nu" (0);
• pentru itemul 2 : "rar" (-1), "uneori"(0), "adesea" (1);
• pentru itemul 3 : "adevărat" (1), "fals" (0);
• pentru itemul 4 : "puţin" (-1), "potrivit" (0), "mult" (1)

atunci

$\mathbf{M}_1=\mathbf{M}_2=\mathbf{M}_3=$ {"da","nu"} x
{"rar","uneori","adesea"} x
{"adevărat", "fals"} x
{"puţin","potrivit","mult"}.

Un element al mulţimii $\mathbf{M}_1$ (un răspuns la prima ocazie de măsurare) poate fi, de pildă, ("da","uneori","fals","mult").

Deci, rezultatele posibile ale experimentelor aleatoare considerate în teoria stărilor şi a trăsăturilor formează produsul cartezian

$$Z = P \times (U_1 \times U_2 \times ... U_n) \times (M_1 \times M_2 \times ... \times M_n)$$

Fiecare element al său, $(p,u_1,u_2,...,u_n,m_1,m_2,...,m_n)\in Z$ este constituit din:

- o persoană $p\in P$;
- n situaţii în care au fost efectuate măsurările (u_k=situaţia apărută la cea de a k-a măsurare, k=1,2,...,n);
- n rezultate de măsurări (m_k=rezultatul măsurării k, k=1,2,...,n, care este format din rezultatele răspunsurilor la itemii componenţi ai chestionarului).

De exemplu, în cazul considerat mai sus, când acelaşi chestionar format din patru itemi a fost administrat de trei ori, un element al mulţimii **Z** poate fi: (Ionescu, "obosit", "bolnav", "bine dispus", "da", "rar", "adevărat", "puţin", "da", "uneori", "adevărat", "puţin", "nu", "rar", "fals", "potrivit")

Deci, persoana "Ionescu", la a doua ocazie de măsurare a fost "bolnavă" şi a răspuns la cei patru itemi ai chestionarului cu "da", "uneori","adevărat" şi "puţin".

Se consideră operatorii:

$p_0 : Z \to P$ care ia ca valori persoanele din **P**;

$p_k : Z \to U_k$ care ia ca valori situaţiile posibile la măsurarea efectuată în ocazia k, k=1,2,...,n.

Se presupune că la fiecare ocazie k, k=1,2,...,n, se măsoară m_k itemi.

În fiecare ocazie de măsurare k, k=1,2,...,n, pentru fiecare item j administrat, j=1,2,...,m_k, se consideră variabila "scor observat", Y_{jk}, care are ca valori scorurile care se pot obţine în populaţia **P**, la itemul j, în cea de a k-a ocazie de măsurare:

$$Y_{jk} : Z \to R$$

În exemplul de mai sus, variabila Y_{32}, asociată celui de-al treilea item al chestionarului la cea de a doua administrare, poate lua valorile 1 (pentru răspunsul "adevărat") şi 0 (pentru răspunsul "fals").

Fiecare variabilă Y_{jk} este o variabilă aleatoare. Ea ia valorile posibile cu anumite probabilităţi.

Se definesc, pentru:

- fiecare persoană $p\in P$;
  - fiecare măsurătoare k, k=1,2,...,n
    - fiecare situaţie posibilă $u_k\in U_k$ în ocazia de măsurare k;
      - fiecare item j, j=1,2,...,m_k

variabilele aleatoare

$$\tau_{jk} = \mathbf{M}(Y_{jk} | p_0{=}p, p_k{=}u_k) \quad (6.3.3.3)$$
$$e_{jk} = Y_{jk} - \tau_{jk} \quad (6.3.3.4)$$
$$\xi_{jk} = \mathbf{M}(Y_{jk} | p_0{=}p) \quad (6.3.3.5)$$
$$\zeta_{jk} = \tau_{jk} - \xi_{jk} \quad (6.3.3.6)$$

τ_{jk} se numeşte "variabila **stare latentă**". Ea reprezintă media tuturor scorurilor pe care le-ar putea obţine persoana *p*, la itemul *j*, în situaţia u_k.

e_{jk} este "eroarea" de măsurare pentru persoana *p*, la itemul *j*, în situaţia u_k. Fiind diferenţa dintre scorul obţinut şi "starea latentă", această variabilă se referă de fapt la eroarea de măsurare a stării latente.

ξ_{jk} se numeşte "variabila **trăsătură latentă**". Ea este media tuturor scorurilor pe care persoana *p* le-ar putea obţine la itemul *j*, la cea de a *k*-a administrare a chestionarului, dacă se iau în considerare toate situaţiile în care s-ar putea afla atunci persoana *p*.

ζ_{jk} este "reziduul stării latente". Ea este diferenţa dintre starea latentă şi trăsătura latentă.

Se observă că toate variabilele (τ_{jk}, ξ_{jk}, ζ_{jk}, e_{jk}) au fost definite relativ la variabila observată Y_{jk}.

Din definiţia variabilei τ_{jk}, (6.3.3.3), rezultă că *o variabilă stare latentă caracterizează o persoană p, într-o situaţie de măsurare* u_k. Fiecare persoană într-o situaţie de măsurare are un scor τ_{jk} şi numai unul.

După cum rezultă din definiţia variabilei ξ_{jk}, (6.3.3.5), *variabila trăsătură latentă caracterizează o persoană p*. Fiecărei persoane îi este atribuit un scor ξ_{jk} şi numai unul.

Definiţia variabilei ζ_{jk}, (6.3.3.6), arată că reziduul stării latente caracterizează persoana *p* într-o situaţie u_k, deoarece fiecărei perechi (p, u_k) îi este atribuit un scor ζ_{jk} şi numai unul.

6.3.4. Modelul de bază al teoriei stărilor şi a trăsăturilor latente

Relaţia (6.3.3.6), transcrisă sub forma

$$\tau_{jk} = \xi_{jk} + \zeta_{jk} \quad (6.3.4.1)$$

arată că pentru orice persoană $p \in P$ şi orice situaţie $u_k \in U_k$, $k=1,2,...,n$, variabila stare latentă τ_{jk} este compusă dintr-o variabilă trăsătură latentă ξ_{jk} şi un reziduu al stării latente, ζ_{jk}.

Relaţia (6.3.3.4) poate fi transcrisă astfel:

$$Y_{jk} = \tau_{jk} + e_{jk} \quad (6.3.4.2)$$

Ecuaţiile (6.3.4.1) şi (6.3.4.2) constituie modelul matematic al teoriei stărilor şi al trăsăturilor latente. El exprimă **ideea de bază** a acestei teorii, şi anume: *atributele manifeste (variabilele observate) au*

- *o componentă eroare*

şi

- *o componentă stare latentă, care constă din*
  - *o componentă trăsătură latentă*

  şi

  - *un rezidu care depinde de situaţia prezentă cu ocazia măsurării şi/sau de interacţiunile dintre persoană şi situaţie.*

6.3.5. Coeficienţii utilizaţi de teoria stărilor şi a trăsăturilor latente

Definiţiile variabilelor latente τ_{jk}, ξ_k, ζ_k şi e_{jk} implică, fără nici o presupunere suplimentară asupra variabilelor, aditivitatea varianţelor corespunzătoare :

$$\sigma^2(Y_{jk}) = \sigma^2(\tau_{jk}) + \sigma^2(e_{jk}) \qquad (6.3.5.1)$$

$$\sigma^2(\tau_{jk}) = \sigma^2(\xi_{jk}) + \sigma^2(\zeta_{jk}) \qquad (6.3.5.2)$$

În ipoteza că variabila Y_{jk} are mai multe valori, deci $\sigma^2(Y_{jk})>0$, se pot defini următorii coeficienţi (Deinzer, Steyer ş.a., 1995):

- **coeficientul de fidelitate:**

$$Rel(Y_{jk}) = \frac{\sigma^2(\tau_{jk})}{\sigma^2(Y_{jk})} \qquad (6.3.5.3)$$

Acesta reprezintă partea din varianţa variabilei observate Y_{jk} datorată variabilei stare latentă. El arată, la fel ca şi în teoria clasică a testelor, cât de precisă este măsurătoarea (în cazul de faţă, cât de precis este măsurată starea latentă).

- **coeficientul de consistenţă:**

$$Con(Y_{jk}) = \frac{\sigma^2(\xi_{jk})}{\sigma^2(Y_{jk})} \qquad (6.3.5.4)$$

Acesta reprezintă partea din varianţa variabilei Y_{jk} datorată numai individului, nu şi situaţiei sau interacţiunii persoanei cu situaţia.

- **coeficientul de specificitate a ocaziei:**

$$Spe(Y_{jk}) = \frac{\sigma^2(\zeta_{jk})}{\sigma^2(Y_{jk})} \qquad (6.3.5.5)$$

El reprezintă partea din varianţa variabilei Y_{jk} datorată situaţiei sau interacţiunii persoanei cu situaţia.

Din relaţia (6.3.5.2) rezultă:

$$\text{Rel}(Y_{jk}) = \text{Con}(Y_{jk}) + \text{Spe}(Y_{jk}) \qquad (6.3.5.6)$$

Deci, un instrument de măsură este precis, dacă valorile înregistrate de el (scorurile observate) surprind atât trăsătura psihică a persoanei, cât şi influenţa situaţiei de măsurare asupra acesteia.

Dacă o variabilă observată Y_{jk} este puternic influenţată de situaţie şi dacă trăsătura persoanei considerate are doar o mică importanţă, atunci coeficientul de specificitate a ocaziei va fi mare şi coeficientul de consistenţă va fi mic. Din contră, un coeficient de consistenţă mare şi un coeficient de specificitate a ocaziei mic arată că variabila observată este influenţată în principal de caracteristicile individului şi că situaţia şi/sau interacţiunea persoanei cu situaţia nu are o importanţă prea mare.

Trebuie remarcat că cei trei coeficienţi definiţi mai sus se referă nu la un test (precum coeficientul de fidelitate în cazul teoriei clasice a testelor), ci la un item de test aplicat unei persoane, într-o anumită situaţie.

6.3.6. Câteva modele ale teoriei stărilor şi a trăsăturilor latente

Ne vom referi în continuare doar la cazul în care se măsoară o singură variabilă trăsătură latentă, prin cel puţin doi itemi, fiecare administrat în două sau mai multe ocazii. Vom prezenta ipotezele utilizate de trei modele (Steyer, Schmitt, 1992).

A. Modelul cu o singură trăsătură (singletrait model) nu face nici o distincţie între stări şi trăsături, adică se presupune că fiecare variabilă stare latentă este identică cu variabila trăsătură latentă care intră în compunerea ei. Deci nu apar efectele ocaziei.

Fiecare variabilă observată Y_{jk} are asociată o variabilă trăsătură latentă, ξ_{jk}. În acest caz, variabilele ξ_{jk} au acelaşi statut ca şi variabilele "scor real" din teoria clasică a testelor. Variabila trăsătură latentă comună, ξ, este definită ca o funcţie liniară de variabilele ξ_{jk}.

B. Modelul cu mai multe stări (multistate model) ia în considerare şi specificitatea situaţiei. Fiecare variabilă observată, Y_{jk} este determinată de două tipuri de influenţe:

- o variabilă stare latentă, comună tuturor situaţiilor de măsurare în ocazia u_k, τ_k;
- o variabilă eroare, e_{jk}.

Acest model nu se ocupă decât de stările latente, nu şi de trăsăturile latente. Fiecărei variabile observate Y_{jk} îi este asociată o variabilă de stare latentă τ_{jk}, iar variabila stare latentă comună τ_k este definită ca o funcţie liniară de variabilele τ_{jk}.

În acest caz, variabilele stare latentă τ_{jk} au acelaşi statut ca şi variabilele "scor real" din teoria clasică a testelor.

C. Modelul cu o singură trăsătură şi mai multe stări (multistate-singletrait model), la fel ca şi modelul cu mai multe stări, nu se interesează de variabila trăsătură latentă, ceea ce înseamnă că nu permite descompunerea varianţei stării latente în varianţa trăsăturii latente şi în varianţa reziduului stării latente.

Se presupune că pentru fiecare ocazie de măsurare u_k există o variabilă stare latentă comună tuturor itemilor, τ_k şi că toate variabilele τ_k au o trăsătură latentă comună, ξ.

CAPITOLUL 7
SURSE DE EROARE ÎN TESTAREA PSIHOLOGICĂ

7.1. NOŢIUNEA DE EROARE

Termenul **eroare** înglobează tot ceea ce scade precizia testului, prin **precizie** înţelegând gradul în care răspunsurile date de subiect sunt predictive pentru o anumită trăsătură ori pentru un anumit comportament (Most, 1987).

Eroarea este rezultatul acţiunii unor factori care influenţează măsurarea într-un asemenea mod încât nu pot fi nici recunoscuţi, nici înţeleşi. Ea afectează scorurile obţinute la test de subiect şi/sau interpretarea dată acestora.

Existenţa erorii (nu şi mărimea ei) este pusă în evidenţă atunci când rezultatele testului sunt folosite pentru a se face predicţii asupra comportamentului şi când se constată că numai o parte dintre predicţii se adeveresc.

7.2. FACTORI CARE INFLUENŢEAZĂ REZULTATELE TESTĂRII PSIHOLOGICE

Scorul obţinut de o persoană la un test psihologic administrat într-o ocazie particulară este rezultatul acţiunii unor factori diverşi. Pe de-o parte, scorul este influenţat de variabila pe care testul trebuie să o măsoare. Aceasta poate fi o caracteristică stabilă a persoanei (o trăsătură) sau o caracteristică dependentă de situaţie (o stare). Pe de altă parte, însă, scorul este determinat şi de numeroşi factori care acţionează ca agenţi perturbatori. Unii dintre ei au legătură cu:

- diverse caracteristici stabile ale persoanei, altele decât cea măsurată de test;
- caracteristicile situaţiei în care are loc testarea;
- caracteristicile testului folosit: tipul itemilor (cu răspunsuri de tip "da"/"nu", cu alegere multiplă etc.), modul de administrare a testului (creion-hârtie, cu ajutorul calculatorului etc.), sarcina impusă de test etc.

iar alţii sunt aleatori. Aceşti factori perturbatori ar putea fi grupaţi astfel (Anastasi, 1976; E. L. Thorndike, după Murphy şi Davidshofer, 1987; Jäger, 1993):

A. ***caracteristici durabile ale persoanei, independente de situaţia de testare şi care au influenţă asupra modului de rezolvare a testelor psihologice în general, nu numai a testului în cauză:***

- abilitatea generală de a înţelege instrucţiunile testelor (persoanele cu un nivel intelectual redus vor avea dificultăţi în a înţelege ce au de făcut);
- deprinderi şi îndemânări generale pentru rezolvarea testelor (persoanele pasionate de rezolvarea diverselor teste întâlnite în reviste se vor descurca mai uşor şi mai rapid cu un test particular);
- unele deprinderi care pot influenţa modul de a rezolva un grup de teste care prezintă asemănări cu testul în cauză (de exemplu, un vânzător care are şi sarcina de a încasa banii pentru marfa vândută, deci de a calcula rapid costul total şi restul, poate fi avantajat la rezolvarea unor teste care conţin şi calcule aritmetice simple, indiferent dacă aceste teste sunt destinate măsurării inteligenţei, a aptitudinii matematice sau a deprinderii de a efectua calcule);

B. ***caracteristici durabile ale persoanei, independente de situaţia de testare şi care au influenţă asupra modului de rezolvare a testului respectiv, dar nu şi a altor teste:***

- cunoştinţe şi deprinderi care pot influenţa modul de rezolvare a testului în cauză (de exemplu, la un test de creativitate care cere să se completeze cât mai multe cuvinte într-un interval de timp dat vor fi avantajate persoanele pasionate de rezolvarea integramelor);
- existenţa unor patterne de răspuns stabile (de exemplu, tendinţa subiectului de a marca prima opţiune, mai des decât altele, într-un test cu alegere multiplă sau tendinţa de a răspunde "adevărat", atunci când nu este sigur care dintre răspunsurile "adevărat" şi "fals" i se potriveşte mai bine);
- cunoaşterea de către subiect a răspunsului la unii itemi (de exemplu, subiectul cunoaşte un fapt particular cerut de un item dintr-un test de cunoştinţe sau ştie la ce se referă un item dintr-un chestionar de personalitate şi cum trebuie să răspundă pentru a crea despre sine imaginea pe care o doreşte);

C. ***factori care acţionează diferit de la o situaţie de testare la alta şi care au influenţă asupra modului de rezolvare de către subiect a mai multor teste psihologic administrate în situaţia respectivă, nu numai a testului în cauză:***

- caracteristici temporare ale persoanei: starea fizică şi psihică generală (starea de sănătate, oboseala, motivaţia, tensiunea emoţională), experienţa anterioară în domeniul testării psihologice, acceptarea necesităţii testării;
- condiţiile fizice în care are loc administrarea testului: căldura, lumina, ventilaţia etc.;

● comportarea examinatorului şi interacţiunea personalităţii, a sexului sau a rasei examinatorului cu subiectul;

D. *factori care acţionează diferit de la o situaţie de testare la alta şi care au influenţă asupra modului de rezolvare de către subiect a testului în cauză, dar nu şi a altor teste care măsoară aceeaşi caracteristică*:

● nivelul de antrenare al subiectului în rezolvarea unor sarcini asemănătoare celei cerute de testul în cauză (de exemplu, la un test de dexteritate manuală care cere parcurgerea unui labirint, o persoană care în zilele precedente testării a efectuat mai multe exerciţii de acest fel are şanse să obţină un punctaj mare);

● dispoziţia momentană a subiectului pentru rezolvarea testului în cauză (de exemplu, unei persoane foarte ocupate în ziua testării i se va părea prea lung un chestionar de personalitate cu mulţi itemi, precum CPI, şi s-ar putea să răspundă la întâmplare ori de câte ori nu este sigură de răspunsul care i se potriveşte);

● fluctuaţii ale memoriei sau ale atenţiei subiectului în timpul testării, care îl fac să nu îşi amintească la moment un anumit fapt sau să nu observe unele amănunte;

E. *factori cu acţiune aleatoare, care afectează administrarea testului în cauză sau aprecierea performanţelor la test, şi care ar putea sau nu să influenţeze rezultatele altor teste administrate în aceeaşi situaţie*:

● factori de distragere a atenţiei subiectului (de exemplu, un zgomot puternic produs în timpul testării);

● norocul la ghicirea răspunsului corect, atunci când subiectul alege răspunsul la întâmplare;

● cotarea greşită a unor itemi, intenţionată sau nu, şi, în consecinţă, obţinerea unui scor eronat.

F. *factori legaţi de concepţia testului în cauză, independenţi de subiect şi de situaţia de examinare*: structura testului, compoziţia şi repartiţia itemilor etc.

Vom analiza, în continuare, câteva dintre cauzele erorilor de măsurare în testările psihologice.

Ambianţa în care are loc administrarea testelor psihologice, chiar dacă nu este apreciată drept nesatisfăcătoare de majoritatea celor examinaţi, pentru unele persoane poate genera nemulţumiri şi poate astfel influenţa rezultatele.

Davidsen şi Maxey (1993) au efectuat un studiu pe un lot de 27631 studenţi, în scopul identificării legăturilor dintre diverse caracteristici de personalitate şi gradul de nemulţumire faţă de câteva condiţii de testare: luminozitatea încăperii, nivelul de zgomot, dimensiunile suprafeţei de scris şi distanţa dintre scaune. Principalele constatări pot fi sintetizate astfel:

- Fiecare dintre aceste condiţii a fost considerată mulţumitoare de cel puţin 89% dintre studenţi, pentru ambele sexe.
- S-a putut observa o creştere a nemulţumirii faţă de toate cele patru condiţii de testare pe măsură ce creşte nivelul de aspiraţie academică.
- În cadrul unor loturi de studenţi cu aceeaşi pregătire şcolară, cei care au cele mai bune rezultate şcolare sunt cei mai nemulţumiţi de mărimea suprafeţei de scris şi de nivelul de zgomot.

Situaţiile de evaluare sunt surse de tensiune şi de anxietate. *Anxietatea faţă de test* este un concept binecunoscut. Se ştie că acest fel de anxietate creşte pe măsură ce creşte importanţa atribuită de subiect evaluării şi cadrului în care se face testarea (Benziman, Toder, 1993). Există studii care arată că dacă anxietatea faţă de test are un nivel uşor mărit, poate creşte motivaţia şi, astfel, se poate îmbunătăţi performanţa, în timp ce un nivel ridicat al anxietăţii faţă de test afectează motivaţia.

Anxietatea este cauzată în mare măsură de ambiguitatea testului, de faptul că subiectul simte că nu poate afla ce se aşteaptă de la el şi de teama că rezultatele testării nu vor fi secrete (Benziman, Toder, 1993).

Lipsa de interes, oboseala sau starea de boală îl pot determina pe subiect să răspundă la întâmplare. O posibilitate de depistare a unor asemenea situaţii, în cazul chestionarelor de personalitate, o constituie introducerea, în test, a unor scale de identificare a inconsistenţei răspunsurilor, de forma celor construite de A. Tellegen (Ozer, Reise, 1994):

- Scala VRIN (variable response inconsistency) constă din perechi de itemi care sunt similari semantic şi la care subiecţii ar trebui să răspundă în acelaşi sens.
- Scala TRIN (true response inconsistency) conţine perechi de itemi care sunt similari semantic dar cotaţi în direcţii opuse. Ea depistează persoanele care răspund la toţi itemii la fel ("adevărat" sau "fals"), fără a fi atente la conţinutul lor.

Gradul de concordanţă a intereselor persoanelor examinate cu cele ale examinatorului are influenţă asupra comportării subiecţilor în timpul testării. Astfel:

• dacă atât subiectul cât şi psihologul sunt interesaţi în stabilirea unui diagnostic corect, este de aşteptat ca motivaţia subiectului să fie ridicată;

• dacă testarea a fost cerută de o autoritate externă, este posibil ca motivaţia subiectului să fie scăzută;

• dacă testarea se face în cadrul unei cercetări, iar subiecţii sunt voluntari, motivaţia poate varia foarte mult de la o persoană la alta (Jäger, 1993).

Motivaţia influenţează atitudinea persoanelor examinate şi, deci, răspunsurile la test.

În cazul utilizării chestionarelor de personalitate, *conţinutul şi forma itemilor testului* pot constitui surse de eroare. Amintim câteva defecte posibile ale itemilor (Angleitner, John, Löhr, 1986; Fife-Schaw, 1995):

a. Itemul nu este înţeles de subiecţi. Aceasta se întâmplă dacă sunt utilizate cuvinte neobişnuite, neologisme, fraze lungi, neclare sau dacă există erori gramaticale.

b. Itemul este ambiguu, el poate fi interpretat în mai multe feluri. Asemenea situaţii apar atunci când itemul conţine cuvinte sau expresii cu mai multe sensuri sau când există incompatibilităţi între item şi formatul răspunsului cerut. De exemplu, mulţi itemi fac referire la ceva care se petrece "frecvent" sau "rar". Or, înţelesul acordat acestor termeni diferă de la o persoană la alta. În urma unui experiment care le-a cerut mai multor persoane să indice cât de des trebuie să apară un eveniment, faţă de numărul total de apariţii posibile, pentru ca ele să îl considere ca fiind "frecvent", s-a constatat că un sfert dintre subiecţi au indicat cu procent mai mic decât 40%, iar un sfert, un procent mai mare decât 80% (Mischel, 1968).

c. Itemul îi cere subiectului să evalueze şi să generalizeze, nu să descrie un comportament particular (de exemplu, în CPI: "L-am privit pe tatăl meu ca fiind omul ideal" sau "Simt o puternică dorinţă de a reuşi în viaţă"). În asemenea situaţii, răspunsurile sunt mai relevante pentru interpretarea dată de subiect itemului decât pentru comportamentul său în situaţii non-test (Mischel, 1968).

d. Itemul cere o informaţie abstractă, de exemplu face referire la o dispoziţie sau o atitudine generală şi nu la un comportament sau la o situaţie concretă. Cu cât itemul este mai abstract, cu atât vor exista mai multe deosebiri între persoanele examinate în privinţa modului în care îl înţeleg.

e. Întrebările sugerează răspunsul. De exemplu, itemul "Sunteţi de acord că politica guvernului în domeniul învăţământului este greşită?" sugerează răspunsul afirmativ.

f. Itemul nu este potrivit populaţiei căreia îi este destinat testul. De exemplu, întrebarea "Câte pahare de bere aţi băut săptămâna trecută ?", adresată

unor persoane cu probleme de sănătate, va primi de la majoritatea subiecţilor răspunsul "0".

g. Itemul conţine două întrebări sau afirmaţii. Persoana pentru care una dintre ele are valoare de adevăr iar cealaltă de fals va răspunde, probabil, la întâmplare. De exemplu, la întrebarea "Consideraţi că programul de instruire pe care l-aţi urmat a fost bun şi eficient pentru dumneavoastră ?" pot să apară două situaţii la care este dificil de dat un răspuns: programul a fost bun, dar ineficient pentru subiect sau programul a fost de slabă calitate, dar pentru persoana examinată a fost eficient.

h. Itemul conţine presupuneri implicite. De exemplu, întrebarea "Cât durează certurile cu soţul/soţia dumneavoastră?" face ipoteza că subiectul este căsătorit şi că se ceartă cu soţul/soţia.

i. Itemul este astfel formulat încât unul dintre răspunsurile posibile îl pun pe subiect într-o lumină negativă. Dorinţa de a face o impresie bună îl va determina să evite acest răspuns, chiar dacă el reflectă o situaţie reală.

j. Prin conţinutul său, itemul deranjează persoanele examinate. Aceasta se întâmplă, de obicei, când întrebările se referă la apartenenţa religioasă sau politică a subiectului.

În mod obişnuit, unui test psihologic i se cere să aibă validitate de aspect, pentru a creşte motivaţia subiecţilor la completarea lui. Frecvent, adulţii refuză să răspundă la un test care li se pare absurd sau ridicol ori răspund la întâmplare. Din acest motiv, *absenţa validităţii de aspect* constituie uneori o sursă de eroare pentru scorul testului.

Alteori, însă, tocmai *prezenţa validităţii de aspect* conduce la denaturarea scorurilor testului. Explicaţia o constituie faptul că testele care au validitate de aspect le permit subiecţilor să ghicească ce măsoară ele. Pentru testele de aptitudini, aceasta nu are importanţă: chiar dacă subiectul ştie că un item măsoară inteligenţa, felul în care va răspunde la el nu se schimbă. Nu acelaşi lucru se întâmplă în cazul testelor de personalitate. De exemplu, la selecţia profesională a poliţiştilor, doar puţine persoane anxioase vor răspunde sincer la un test care măsoară anxietatea şi are validitate de aspect.

CAPITOLUL 8
FIDELITATEA TESTELOR PSIHOLOGICE

8.1 DEFINIŢIA FIDELITĂŢII TESTELOR

În vorbirea curentă, cuvântul "fidel" este folosit pentru a descrie constanţa (sentimentelor), durabilitatea (ataşamentului), exactitatea (memoriei) sau, în cazul unui instrument de măsură, proprietatea acestuia de a arăta aceeaşi valoare când se repetă măsurarea (Le petit Larousse, 1993).

În teoria testelor, conceptul de fidelitate a cunoscut diverse accepţiuni pe parcursul timpului. Dacă în 1954 şi în 1966, *Standards for Educational and Psychological Tests* înţelegeau prin fidelitate "precizia -consistenţa şi stabilitatea- măsurării realizate de test", în 1974 fidelitatea este privită ca şi "gradul în care rezultatele testării pot fi atribuite surselor sistematice de variaţie" (*Standards for Educational and Psychological Tests*, 1974).

Termenii de "consistenţă" şi "stabilitate" continuă să fie utilizaţi, în legătură cu fidelitatea testelor, având sensul de repetabilitate a rezultatelor măsurării. Astfel, ultimele *Standards for Educational and Psychological Tests*, cele din 1985, definesc fidelitatea ca fiind "gradul în care scorurile testului sunt consistente sau repetabile, adică gradul în care ele nu sunt afectate de erorile de măsură".

În unele definiţii, prin fidelitate se înţelege "precizia măsurării" realizate de un instrument. Trebuie menţionat, însă, că termenul de "precizie" (accuracy) apare în literatura psihologică în diverse accepţiuni. De exemplu (Silva, 1993):

- R. G. Lord îl utilizează ca sinonim pentru "fidelitate";
- I. Delclaux şi M. R. Martínez Arias consideră că el se referă la mărimea erorii standard de măsură;
- J. M. Johnston şi H. S. Pennypacker îl folosesc pentru a desemna gradul în care valorile obţinute prin măsurare aproximează starea reală din natură;
- A. E. Kazdin îl interpretează ca fiind gradul în care o înregistrare făcută cu un instrument de măsură se potriveşte cu mărimea unui criteriu prestabilit referitor la aceleaşi date.

În continuare vom urmări definiţia fidelităţii în termeni matematici, în teoria clasică a testelor.

Se notează cu **M** o mulţime de teste şi cu **P** o mulţime de persoane. Se alege un test $m \in \mathbf{M}$, care se aplică tuturor persoanelor din **P**.

Conform definiţiei formulate în paragraful 6.1.1, **coeficientul de fidelitate al testului** m este pătratul coeficientului de corelaţie liniară între scorurile reale şi scorurile observate la test, $\rho^2(X_{m*},T_{m*})$. El verifică relaţia 6.1.1.23:

$$\rho^2(X_{m*},T_{m*}) = \frac{\sigma^2(T_{m*})}{\sigma^2(X_{m*})}$$

Fiind pătratul unui coeficient de corelaţie, coeficientul de fidelitate este cuprins întotdeauna între 0 şi 1.

Formula de definiţie a coeficientului de fidelitate se poate rescrie, folosind relaţia 6.1.1.22, astfel:

$$\rho^2(X_{m*},T_{m*}) = 1 - \frac{\sigma^2(E_{m*})}{\sigma^2(X_{m*})} \qquad (8.1.1)$$

De aici se deduc următoarele:

A. Coeficientul de fidelitate este egal cu 1 dacă şi numai dacă eroarea standard de măsură $\sigma_E=\sigma(E_{m*})$ este nulă. Deoarece media erorilor de măsură $M(E_{m*})$ este nulă, dispersia erorilor de măsură poate fi egală cu zero numai dacă toate erorile de măsură sunt nule. Deci, un test are coeficientul de fidelitate egal cu unu dacă şi numai dacă scorurile observate coincid cu scorurile reale, adică dacă testul este o măsură foarte precisă.

B. Coeficientul de fidelitate este egal cu zero dacă şi numai dacă are loc $\sigma^2(T_{m*})=0$, adică toate persoanele din populaţia P au acelaşi scor real.

Dacă $a\in$M şi $b\in$M sunt două instrumente de măsură paralele şi dacă pentru orice persoană $p\in$P scorurile observate X_{ap} şi X_{bp} (sau, echivalent, erorile de măsură E_{ap} şi E_{bp}) sunt variabile aleatoare independente, atunci, conform teoremei 6.1.2.3, coeficientul de fidelitate poate fi calculat cu formula 6.1.2.13:

$$\rho^2(X_{a*},T_{a*}) = \rho(X_{a*},X_{b*})$$

De aici rezultă că un test poate avea coeficientul de fidelitate egal cu unu dacă şi numai dacă $X_{a*}=X_{b*}$.

În concluzie, putem spune, că un test este fidel atunci când:

• este precis (lipsit de erori de măsură)

sau, echivalent,

• orice persoană obţine scoruri egale la test şi la o formă paralelă a acestuia, condiţiile de administrare fiind stabilite astfel încât erorile de măsură la cele două teste, pentru orice persoană $p \in P$, să fie independente.

În studiul fidelităţii se presupune că erorile de măsură sunt aleatoare şi nu sistematice. O eroare este sistematică dacă afectează în acelaşi mod valoarea observată, la fiecare măsurare. De exemplu, dacă pentru aprecierea performanţelor unor operatori se numără greşelile efectuate de aceştia la introducerea unui text în calculator şi una dintre taste este defectă, atunci caracterul respectiv va apărea pentru toţi subiecţii ca fiind tastat greşit. Această eroare va fi prezentă în toate măsurătorile efectuate. Ea este o eroare sistematică.

Teoria referitoare la fidelitatea testelor consideră că erorile de măsură sunt complet impredictibile, indiferent de cunoştinţele pe care le avem despre persoanele măsurate sau despre procesul de măsurare. Înainte de a calcula coeficientul de fidelitate al unui test este absolut necesar să se precizeze ce reprezintă scorurile reale ale testului. Aceasta, deoarece o anumită caracteristică a unei persoane poate fi măsurată prin scorurile reale ale unui test sau poate influenţa scorurile observate, ca eroare de măsură, la un alt test.

Să considerăm un test de cunoştinţe de fizică administrat cu ajutorul calculatorului. În acest caz, lipsa deprinderii de a lucra la calculator şi anxietatea faţă de calculator acţionează ca şi factori perturbatori, mărind erorile de măsură. Influenţa lor va fi mai puternică la elevii unei clase cu profil umanist decât la cei dintr-o clasă de informatică. Dacă acelaşi test este administrat în varianta creion-hârtie, factorii menţionaţi nu vor mai avea nici un rol.

Se deduce, de aici, că pentru a arăta cât de bine sunt reflectate scorurile reale ale unui test de către scorurile observate nu este suficient să se indice o valoare a coeficientului de fidelitate, ci trebuie să se precizeze:

• caracteristicile persoanelor care au răspuns la test (nivel de studii, vârstă etc.);

• condiţiile în care a fost administrat testul.

Deci, nu se vorbeşte despre fidelitatea unui instrument de măsură, ci despre fidelitatea acelui test aplicat la o anumită populaţie şi în anumite condiţii.

8.2. ROLUL COEFICIENTULUI DE FIDELITATE

Coeficientul de fidelitate serveşte la aprecierea preciziei unui test, dar nu este suficient pentru descrierea acestuia ca instrument de măsură. O aceeaşi

valoare a coeficientului de fidelitate poate fi obţinută pentru dispersii diferite ale scorurilor observate şi ale erorilor de măsură. Coeficientul de fidelitate reflectă doar mărimea raportului celor două dispersii. Din această cauză, *Standards for Educational and Psychological Tests* din 1985 recomandă ca în manualul testului să figureze, alături de coeficientul de fidelitate al testului, eroarea standard de măsură, σ_E.

Coeficienţii de fidelitate se utilizează la compararea testelor. Eroarea standard de măsură serveşte la interpretarea scorurilor testului, mai precis, la determinarea intervalelor de încredere pentru scorurile reale ale subiecţilor.

Raţionamentul care stă la baza procedurii de estimare a intervalului de încredere pentru scorul real al unui subiect este identic cu cel utilizat în statistică pentru estimarea intervalului de încredere al mediei populaţiei (Traub, 1994). Se presupune că valorile variabilei X_{mp} pentru persoana $p \in \mathbf{P}$, la testul $m \in \mathbf{M}$, sunt repartizate normal, cu media τ_{mp} şi cu abaterea standard $\sigma(E_{mp})$. Cunoscând valoarea unui scor observat pentru persoana p, x_{mp} şi considerând valoarea $\sigma(E_{mp})$ egală cu eroarea standard de măsură în populaţia $\mathbf{P}$, ${}_E\sigma$, intervalul de încredere al scorului real, corespunzător unei probabilităţi de 0.68, este

$$(x_{mp}-\sigma_E, x_{mp}+\sigma_E),$$

iar cel corespunzător unei probabilităţi de 0.95 este

$$(x_{mp}-1.96\sigma_E, x_{mp}+1.96\sigma_E)$$

Eroarea standard se calculează cu ajutorul coeficientului de fidelitate şi al abaterii standard a scorurilor observate:

$$\sigma_E = \sigma(X_{m*})\sqrt{1-\rho^2(X_{m*},T_{m*})} \quad (8.2.1)$$

În raportul psihologic întocmit pentru persoana p, alături de scorul obţinut de aceasta la test, x_{mp}, se va indica şi intervalul de încredere al scorului real.

O altă utilizare a coeficienţilor de fidelitate o constituie calculul coeficientului de corelaţie liniară între scorurile reale ale subiecţilor la două teste, în scopul studierii asemănării constructelor măsurate de acestea. Notând cu a şi b cele două teste, se aplică următoarea formulă (Perugini, Gallucci, 1979):

$$\rho(T_{a*}, T_{b*}) = \frac{\rho(X_{a*}, X_{b*})}{\sqrt{\rho^2(X_{a*}, T_{a*})\rho^2(X_{b*}, T_{b*})}} \quad (8.2.2)$$

unde:

$\rho(T_{a^*}, T_{b^*})$ este coeficientul de corelaţie liniară între scorurile reale la cele două teste;

$\rho(X_{a^*}, X_{b^*})$ este coeficientul de corelaţie liniară între scorurile observate la cele două teste;

$\rho^2(X_{a^*}, T_{a^*})$ şi

$\rho^2(X_{a^*}, T_{a^*})$ sunt coeficienţii de fidelitate ai testelor.

Din această formulă se observă că totdeauna coeficientul de corelaţie liniară între scorurile observate la teste, în valoare absolută, este mai mic sau cel mult egal cu valoarea absolută a coeficientului de corelaţie liniară între scorurile reale ale testelor.

8.3. FORMULE DE CALCUL UTILIZATE PENTRU EVALUAREA FIDELITĂŢII

Întrucât repartiţiile variabilelor aleatoare X_{m^*}, T_{m^*} şi E_{m^*} pentru un test *m* nu pot fi cunoscute, coeficientul său de fidelitate nu poate fi calculat pe baza formulei de definiţie (6.1.1.23). Unele metode de testare psihologică (metoda test-retest, metoda formelor paralele) aplică formula 6.1.2.13. Dar aceasta nu furnizează decât o valoare aproximativă, deoarece este imposibil de stabilit dacă două teste sunt cu adevărat paralele. În această situaţie, se obişnuieşte să se estimeze coeficientul de fidelitate folosindu-se diverse formule care, în anumite condiţii, acceptabile din punct de vedere teoretic (dar a căror îndeplinire nu poate fi probată în practică), furnizează valori egale cu coeficientul de fidelitate.

Considerăm că un coeficient este convenabil pentru estimarea fidelităţii unui test psihologic dacă îndeplineşte următoarele condiţii (Albu, 1990):

a. este cuprins în intervalul [0,1], la fel ca şi coeficientul de fidelitate;

b. are o valoare apropiată de cea a coeficientului de fidelitate;

c. are o comportare consecventă faţă de coeficientul de fidelitate, adică este permanent mai mare sau permanent mai mic decât acesta;

d. poate fi calculat cu uşurinţă, fără a pretinde experimente suplimentare.

Vom urmări aceste aspecte la câţiva coeficienţi care se utilizează în practică pentru testele formate din *n* părţi (itemi sau grupe de itemi), $n \geq 2$, atunci când scorul observat al testului se obţine prin însumarea scorurilor observate ale părţilor componente:

$$X = Y_1 + Y_2 + \ldots + Y_n$$

unde

X este variabila aleatoare corespunzătoare scorurilor observate la întregul test;

Y_i este variabila aleatoare corespunzătoare scorurilor observate la cea de-a i-a parte a testului, $i=1,2,\ldots,n$.

Aceşti coeficienţi sunt:

- coeficientul α al lui Cronbach (Cronbach, 1943):

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum_{i=1}^{n}\sigma_i^2}{\sigma_X^2}\right) \qquad (8.3.1)$$

- coeficientul λ_3 al lui Guttman (Lord, Novick, 1968):

În ambele formule:

$$\lambda_3 = 1 - \frac{\sum_{i=1}^{n}\sigma_i^2}{\sigma_X^2} + \frac{1}{\sigma_X^2}\sqrt{\frac{n}{n-1}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n}\sigma^2(Y_i, Y_j)} \qquad (8.3.2)$$

$\sigma_X^2 = \sigma^2(X)$ este dispersia scorurilor observate ale testului;

$\sigma_i^2 = \sigma^2(Y_i)$ reprezintă dispersia scorurilor observate la cea de-a i-a parte a testului, pentru $i=1,2,\ldots,n$;

$\sigma(Y_i, Y_j)$ este covarianţa scorurilor observate la partea a i-a şi la partea a j-a a testului, $i,j=1,2,\ldots,n$ şi $i\neq j$.

- coeficienţii ρ_{20} şi ρ_{21} ai lui Kuder şi Richardson, utilizabili atunci când cele n componente ale testului sunt itemi binari, adică Y_i, $i=1,2,\ldots,n$, sunt variabile aleatoare cu repartiţii de forma:

$$Y_i : \begin{pmatrix} 1 & 0 \\ p_i & q_i \end{pmatrix}, \quad p_i + q_i = 1$$

(Richardson, Kuder, 1939):

$$\rho_{20} = \frac{n}{n-1}\left(1 - \frac{\sum_{i=1}^{n} p_i q_i}{\sigma_X^2}\right) \qquad (8.3.3)$$

$$\rho_{21} = \frac{n}{n-1}\left(1 - \frac{n\overline{pq}}{\sigma_X^2}\right) \qquad (8.3.4)$$

Observaţie. Coeficientul ρ_{20} corespunde coeficientului α: în cazul itemilor binari formula 8.3.1 devine formula 8.3.3.

În formula 8.3.4 s-au folosit notaţiile:

$$\bar{p} = \frac{1}{n}\sum_{i=1}^{n} p_i \qquad (8.3.5)$$

$$\bar{q} = \frac{1}{n}\sum_{i=1}^{n} q_i \qquad (8.3.6)$$

- coeficientul α_H al lui Horst, obţinut prin modificarea coeficientului ρ_{20} (Guilford, 1965):

$$\alpha_H = \frac{\sigma_X^2 - \sum_{i=1}^{n} p_i q_i}{\sigma_m^2 - \sum_{i=1}^{n} p_i q_i} * \frac{\sigma_m^2}{\sigma_X^2} \qquad (8.3.7)$$

unde:

$$\sigma_m^2 = 2\sum_{i=1}^{n} R_i p_i - M_X(1+M_X) \qquad (8.3.8)$$

R_i = rangul itemului i din test, itemii fiind ordonaţi după dificultate, cel mai uşor având rangul n. R_i este număr natural.

M_X = media scorurilor observate la test.

Să urmărim proprietăţile acestor coeficienţi.

Vom nota în continuare cu δ diferenţa dintre dispersia scorurilor observate la test şi suma dispersiilor scorurilor observate ale componentelor:

$$\delta = \sigma_X^2 - \sum_{i=1}^{n} \sigma_i^2 \qquad (8.3.9)$$

şi cu

$\rho^2_{XT} = \rho^2(X_{m^*}, T_{m^*})$

coeficientul de fidelitate al testului.

TEOREMA 8.3.1 (Albu, 1990). Dacă pentru orice persoană $p \in P$ scorurile observate la oricare două părţi ale testului, Y_{ip} şi Y_{jp} sunt variabile aleatoare independente, atunci au loc inegalităţile:

$$\rho^2_{XT} \geq \lambda_3 \geq \alpha \qquad (8.3.10)$$

λ_3 este egal cu α dacă şi numai dacă toate covarianţele $\sigma(Y_i, Y_j)$ sunt egale între ele, $i,j=1,2,...,n$ şi $i \neq j$.

Deci λ_3 şi α sunt valori în general mai mici decât coeficientul de fidelitate, dar λ_3 este mai apropiat de ρ^2_{XT} decât α.

TEOREMA 8.3.2 (Albu, 1990).Dacă testul este compus din n părţi care constituie măsurători paralele, atunci au loc egalităţile:

$$\rho^2_{XT} = \lambda_3 = \alpha \qquad (8.3.11)$$

În plus, în acest caz coeficientul de fidelitate ρ^2_{XT} este calculat cu următoarea formulă, numită formula generală Spearman-Brown pentru fidelitatea unui test care are componente paralele:

$$\rho^2_{XT} = \frac{n\rho(Y_1,Y_2)}{1+(n-1)\rho(Y_1,Y_2)} \qquad (8.3.12)$$

Observaţii

A. Componentele testului fiind paralele, au loc relaţiile

$$\rho(Y_i,Y_j) = \rho(Y_1,Y_2),\ \ i,j=1,2,...,n,\ i\neq j \qquad (8.3.13)$$

şi deci fiecare componentă va avea acelaşi coeficient de fidelitate, egal cu $\rho(Y_1,Y_2)$.

B. Se observă că:

$$\rho^2_{XT} = \rho(Y_1,Y_2) \text{ dacă } \rho(Y_1,Y_2)=0 \text{ sau } \rho(Y_1,Y_2)=1$$

şi

$$\rho^2_{XT} > \rho(Y_1,Y_2) \text{ dacă } \rho(Y_1,Y_2)\in(0,1)$$

Adică fidelitatea unui test format din părţi paralele este cel puţin egală cu fidelitatea unei părţi.

C. Formula lui Spearman-Brown a fost generalizată pentru cazul când scorul observat al testului se calculează ca sumă ponderată a scorurilor observate ale părţilor componente, astfel (Li, Rosenthal, Rubin, 1996):

$$\rho^2_{XT} = 1 - \frac{\sum_{i=1}^{n} w_i(1-r_i)\sigma^2(Y_i)}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma(Y_i,Y_j)} \qquad (8.3.14)$$

unde:

w_i = ponderea atribuită părţii a i-a a testului, $i=1,2,...,n$;

r_i = coeficientul de fidelitate al părţii a i-a a testului, $i=1,2,...,n$.

Dacă toate ponderile w_i, $i=1,2,...,n$, au valoarea 1 şi dacă cele n părţi ale testului constituie măsurători paralele, atunci coeficienţii de fidelitate ai tuturor părţilor iau aceeaşi valoare şi formula (8.3.14) coincide cu formula lui Spearman-Brown.

TEOREMA 8.3.3 (Albu, 1990). În general, au loc inegalităţile:

$$\rho^2_{XT} \geq \rho_{20} \geq \rho_{21} \qquad (8.3.15)$$

Coeficienţii ρ_{20} şi ρ_{21} coincid dacă şi numai dacă toate probabilităţile p_i, $i=1,2,...,n$, sunt egale, adică toate variabilele aleatoare Y_i au aceeaşi repartiţie.

Deci valoarea furnizată de ρ_{20} aproximează mai bine coeficientul de fidelitate decât ρ_{21}.

TEOREMA 8.3.4 (Albu, 1990). Dacă cele n părţi ale testului reprezintă măsurători paralele, atunci are loc egalitatea:

$$\rho^2_{XT} = \alpha_H \qquad (8.3.16)$$

Nici unul dintre coeficienţii prezentaţi mai sus (α, λ_3, ρ_{20}, ρ_{21}, α_H) nu depăşeşte valoarea 1.

În ceea ce priveşte semnul acestor coeficienţi, ei pot lua atât valori pozitive, cât şi valori negative.

TEOREMA 8.3.5 (Albu, 1990). Au loc următoarele egalităţi:

$$\text{sgn}\ \alpha = \text{sgn}\ \alpha_H = \text{sgn}\ \rho_{20} = \text{sgn}\ \delta \qquad (8.3.17)$$

Atunci când δ este nenegativ, şi coeficientul λ_3 este nenegativ; în schimb, dacă δ este negativ, coeficientul λ_3 poate fi pozitiv, negativ sau nul.

Întrucât coeficientul de fidelitate al unui test este un număr nenegativ, valorile coeficienţilor folosiţi pentru estimarea sa sunt convenabile doar dacă sunt pozitive.

Din cele arătate mai sus rezultă că cel mai util coeficient pentru estimarea fidelităţii unui test este λ_3, pentru că:

a. poate fi calculat şi pentru teste ai căror itemi nu sunt binari;

b. este mai apropiat de coeficientul de fidelitate decât α;

c. are o comportare consecventă faţă de coeficientul de fidelitate, fiind totdeauna cel mult egal cu acesta ($\lambda_3 \le \rho^2_{XT}$);

d. poate lua valori pozitive şi în unele cazuri când ceilalţi coeficienţi sunt negativi.

Să analizăm în continuare cazul particular n=2. Deci testul este alcătuit din două părţi.

Dacă aceste componente sunt măsurători paralele, atunci $\rho(Y_1,Y_2)$ reprezintă coeficientul de fidelitate al fiecăreia dintre părţi (Y_1 sau Y_2) şi coeficientul de fidelitate al întregului test poate fi calculat cu ajutorul formulei lui Spearman-Brown:

$$\rho^2_{XT} = \frac{2\rho(Y_1,Y_2)}{1+\rho(Y_1,Y_2)} \qquad (8.3.18)$$

Spearman şi Brown au dedus această formulă impunând părţilor testului condiţii mai slabe decât cele cerute de paralelism. Astfel, s-a presupus că se

poate construi un test alcătuit din două părţi, cu variabilele aleatoare asociate scorurilor observate Y_3 şi Y_4, care să verifice egalităţile:

$$\sigma(Y_1) = \sigma(Y_2) = \sigma(Y_3) = \sigma(Y_4) \qquad (8.3.19)$$

şi

$$\rho(Y_i,Y_j) = \text{constant},\ i,j=1,2,3,4,\ i\neq j \qquad (8.3.20)$$

În aceste condiţii are loc relaţia :

$$\rho(Y_1+Y_2,Y_3+Y_4) = \frac{2\rho(Y_1,Y_2)}{1+\rho(Y_1,Y_2)} \qquad (8.3.21)$$

unde $\rho(Y_1+Y_2, Y_3+Y_4)$ este tocmai coeficientul de fidelitate al testului.

Se constată că pentru a putea aplica formula lui Spearman- Brown trebuie să se verifice doar egalitatea $\sigma(Y_1)=\sigma(Y_2)$ şi să se presupună că se poate construi un test ale cărui părţi Y_3 şi Y_4 să satisfacă cerinţele formulate prin egalităţile 8.3.19 şi 8.3.20. Practic este, însă, foarte greu de realizat un asemenea test.

Flanagan (1955) a definit echivalenţa a două teste, X şi Z, fiecare alcătuit din două părţi, unul având scorurile observate $X=Y_1+Y_2$ iar celălalt cu scorurile observate $Z=Y_3+Y_4$, prin condiţiile:

$$\sigma(Y_1+Y_2) = \sigma(Y_3+Y_4) \qquad (8.3.22)$$

(scorurile observate la cele două teste au aceeaşi dispersie) şi

$$\sigma(Y_i,Y_j) = \text{constant, pentru } i,j=1,2,3,4,\quad i\neq j \qquad (8.3.23)$$

(covarianţa oricărei perechi de părţi de test este aceeaşi)

Dacă sunt îndeplinite relaţiile 8.3.21 şi 8.3.22, atunci coeficientul de fidelitate al oricăruia dintre testele X şi Z poate fi calculat cu formula coeficientului α al lui Cronbach:

$$\rho(X,Z) = 2\left(1 - \frac{\sigma^2(Y_1)+\sigma^2(Y_2)}{\sigma^2(Y_1+Y_2)}\right) \qquad (8.3.24)$$

Se constată că în această formulă nu intervin valori statistice ale variabilelor aleatoare Y_3 şi Y_4. Deci nici nu trebuie să existe testul cu scorurile observate $Z=Y_3+Y_4$. Dar, este deosebit de dificil ca, având un test alcătuit din două părţi, să se poată şti dacă se poate construi un test echivalent cu acesta în sensul lui Flanagan.

De remarcat că, dacă două teste îndeplinesc condiţiile impuse de Spearman şi Brown, atunci ele sunt echivalente în sensul lui Flanagan, iar formula 8.3.24 se reduce la formula lui Spearman-Brown.

În cazul n=2 coeficientul λ_3 devine:

$$\lambda_3 = \begin{cases} 2\left(1 - \dfrac{\sigma^2(Y_1)+\sigma^2(Y_2)}{\sigma^2(Y_1+Y_2)}\right), & daca\ \rho(Y_1,Y \\ 0 \qquad\qquad\qquad\qquad , & daca\ \rho(Y_1,Y \end{cases} \tag{8.3.25}$$

Deci, dacă $\rho(Y_1,Y_2) \geq 0$, atunci λ_3 se calculează prin formula 8.3.24.

8.4. METODE PENTRU EVALUAREA FIDELITĂŢII TESTELOR

Din cele arătate până acum se constată că în general nu se poate cunoaşte valoarea exactă a unui coeficient de fidelitate. Fiecare din formulele de calcul prezentate furnizează câte o valoare care coincide cu coeficientul de fidelitate numai în anumite condiţii particulare, greu de realizat.

Cu toate acestea, pentru un test, în funcţie de ceea ce măsoară şi de condiţiile în care a fost administrat, se calculează unul sau mai mulţi dintre următorii coeficienţi, prin care se aproximează valoarea coeficientului de fidelitate:

A. *coeficienţii de stabilitate* - se calculează prin metoda test- retest, corelând scorurile obţinute la test şi la retest (acelaşi test aplicat după un interval de timp). Arată cât de stabile sunt scorurile în timp.

B. *coeficienţii de echivalenţă* - rezultă din metoda formelor paralele, prin corelarea scorurilor obţinute la două teste paralele, administrate aproape în acelaşi timp. Arată cât de asemănătoare sunt cele două instrumente de măsură.

C. *coeficienţii consistenţei interne* - se calculează în cadrul analizei consistenţei interne, pe baza scorurilor obţinute la un test administrat o singură dată. Din această categorie fac parte coeficientul α al lui Cronbach, coeficientul λ_3 al lui Guttman şi coeficienţii Kuder- Richardson. Ei arată concordanţa diferitelor părţi ale testului.

D. *coeficienţii de fidelitate interevaluatori* - se calculează pentru teste al căror scor este rezultatul aprecierii subiective făcute de evaluator. Ei arată în ce măsură părerile mai multor evaluatori concordă între ele.

Prezentăm în continuare aceste tipuri de coeficienţi, metodele prin care se obţin şi condiţiile în care pot fi utilizaţi.

8.4.1. Metoda test-retest

Metoda test-retest evaluează gradul în care scorurile obţinute la un test de acelaşi subiect sunt constante de la o administrare la alta.

Se procedează astfel:

(1) Se administrează testul unui grup de persoane.

(2) După un interval de timp se administrează testul, aceloraşi indivizi, în aceleaşi condiţii ca şi prima dată.

(3) Se calculează coeficientul de corelaţie liniară între scorurile observate în cele două situaţii.

Valoarea obţinută se foloseşte pentru estimarea fidelităţii testului, considerându-se că testul este paralel cu el însuşi, ceea ce înseamnă că între cele două administrări ale sale, scorurile reale ale persoanelor nu s-au schimbat.

Coeficientul de corelaţie calculat prin această metodă se numeşte **coeficient de stabilitate.**

Metoda test-retest este utilă atunci când scorurile reale ale testului măsoară caracteristici *durabile*, generale şi specifice, ale persoanelor.

Temporar	Durabil	
	xxxxxxxx	General
	xxxxxxxx	Specific

Factorii temporari (starea de oboseală sau boală a celui examinat, emoţiile, temperatura şi lumina din sala de testare etc.) pot ajuta unei persoane într-o ocazie şi îi pot scădea scorul observat la un alt experiment. Ei nu modifică scorul real al persoanei, ci influenţează erorile de măsură.

Aplicarea metodei test-retest impune să se asigure asemenea condiţii încât pentru fiecare persoană erorile de măsură la cele două testări să fie variabile aleatoare independente. Aceasta înseamnă, de exemplu, că intervalul de timp dintre test şi retest trebuie stabilit astfel încât subiecţii să nu îşi poată

aminti cum au răspuns la test, întrucât multe persoane sunt tentate să repete, la întrebările la care nu sunt sigure de răspunsul corect sau potrivit, răspunsurile pe care le-au dat, la întâmplare, cu ocazia primei testări. Se recomandă ca lungimea acestui interval să nu fie mai scurtă de trei luni (Kline, 1993).

Adesea, mai ales în cazul testelor de aptitudini şi de cunoştinţe, în intervalul de timp dintre test şi retest subiecţii îşi modifică scorul real, datorită maturizării, a experienţei sau a cunoştinţelor dobândite.

Să analizăm, de exemplu, ceea ce se poate întâmpla după prima administrare a unui test de cunoştinţe. Pe de o parte, intervine uitarea, care determină o scădere a scorurilor reale. Pe de altă parte, subiecţii pot dobândi cunoştinţe noi, ceea ce face să crească scorurile reale. Unele persoane pot căuta în cărţi răspunsurile corecte la itemi, mai ales acolo unde nu au ştiut sau nu au fost sigure de răspuns. Şi astfel, nivelul lor de cunoştinţe, deci scorul real, creşte.

În asemenea situaţii, coeficientul de corelaţie liniară între scorurile observate la test şi la retest nu măsoară fidelitatea testului de cunoştinţe. El arată cum se schimbă nivelul cunoştinţelor verificate la persoanele examinate. Pentru o mai bună interpretare a modificărilor intervenite în persoane, între test şi retest, este util să se calculeze media diferenţelor de scor dintre cele două administrări ale testului şi să se verifice dacă aceasta diferă semnificativ de zero.

Valoarea coeficientului de corelaţie liniară test-retest poate fi apropiată de 1 şi atunci când testul nu este potrivit lotului de subiecţi, fiind foarte dificil sau, din contră, foarte uşor. În asemenea situaţii, majoritatea persoanelor vor obţine, la ambele administrări ale testului, scoruri foarte mici, respectiv foarte mari. Prin urmare, interpretarea valorii coeficientului de stabilitate al unui test trebuie însoţită de analiza formei histogramelor scorurilor observate la test şi la retest.

Totdeauna, împreună cu valoarea unui coeficient de stabilitate se vor comunica:

- lungimea intervalului de timp dintre test şi retest;
- condiţiile în care s-au efectuat testările, pentru a putea identifica eventualele surse de eroare de măsură;
- vârsta subiecţilor testaţi, întrucât s-a constatat că şi aceasta are influenţă asupra mărimii coeficientului de stabilitate (Schuerger, Zarrella, Hotz, 1989).

O altă formulă pentru calculul coeficientului de stabilitate se aplică atunci când testul este utilizat pentru a se lua decizii, prin plasarea subiecţilor în una din categoriile "admis" (cei cu scor mai mare decât un prag fixat) sau "respins" (cei cu scor mai mic decât acest prag). Testul este stabil dacă de la o administrare la alta nu se produc modificări ale clasificării subiecţilor.

Rezultatele clasificărilor realizate pe baza testului şi a retestului pot fi reprezentate sub forma următorului tabel de frecvenţe:

Test	Retest	
	respinşi	admişi
respinşi	a	b
admişi	c	d

Notând cu N numărul total al subiecţilor care au participat la ambele administrări ale testului ($N=a+b+c+d$), coeficientul de stabilitate se calculează cu următoarea formulă (Aiken, 1994):

$$s=\frac{ad-bc}{ad-bc+N*\min\{b,c\}} \qquad (8.4.1.1)$$

Trebuie remarcat că acest coeficient ia valoarea maximă, egală cu 1, nu numai atunci când clasificarea realizată pe baza testului coincide cu cea rezultată în urma retestării (când $b=c=0$), ci şi dacă una singură dintre frecvenţele b şi c este nulă, frecvenţele a şi d fiind, ambele, nenule.

Dezavantajul metodei test-retest constă în faptul că pretinde două administrări ale testului, ceea ce necesită timp şi cheltuieli materiale.

8.4.2 Metoda analizei consistenţei interne

Această metodă utilizează pentru estimarea fidelităţii unui test dispersiile şi covarianţele scorurilor observate ale itemilor. Este avantajoasă, necesitând o singură administrare a testului. Pe baza scorurilor observate ale itemilor se calculează de obicei unul dintre următorii coeficienţi:

- coeficientul α al lui Cronbach;
- coeficientul λ_3 al lui Guttman;
- coeficientul ρ_{20} sau ρ_{21} al lui Kuder-Richardson (dacă itemii testului sunt binari).

Aceşti coeficienţi sunt denumiţi **coeficienţi de consistenţă internă**. Ei indică măsura în care itemii testului se referă la acelaşi lucru.

Faptul că, în general, coeficienţii de consistenţă internă coincid cu coeficientul de fidelitate al testului atunci când itemii componenţi sunt paraleli,

adică măsoară acelaşi lucru şi îl măsoară la fel de bine, conduce la concluzia că aceşti coeficienţi sunt utili pentru calculul fidelităţii testelor care măsoară o caracteristică unidimensională.

Analizând formulele lor de calcul, se constată că valorile coeficienţilor de consistenţă internă cresc atunci când intercorelaţiile itemilor, $\rho(Y_i,Y_j)$, $i,j=1,2,...,n$, $i\neq j$, sunt mari. În schimb, dacă toate intercorelaţiile $\rho(Y_i,Y_j)$, $i\neq j$ sunt egale cu zero, adică itemii testului nu sunt corelaţi liniar între ei, atunci coeficienţii α, λ_3 şi ρ_{20} sunt egali cu zero, iar coeficientul ρ_{21} este negativ.

Se impune o observaţie: dacă testul este administrat cu limită de timp, atunci ultimii itemi nu vor fi rezolvaţi de mai multe persoane, deci în lotul de subiecţi ei vor apărea de mai multe ori cu scorul zero. Intercorelaţiile lor vor avea, din această cauză, valori mari, crescând astfel consistenţa internă a testului. În concluzie, pentru a putea calcula coeficienţii de consistenţă internă ai unui test, acesta trebuie administrat fără limită de timp.

În cazul testelor compuse din itemi binari, consistenţa internă este mare atunci când itemii au aproximativ aceeaşi dificultate şi majoritatea persoanelor care rezolvă corect un item îl rezolvă corect şi pe celălalt.

O altă legătură între caracteristicile itemilor şi fidelitatea testului se obţine rescriind formula coeficientului α al lui Cronbach, cu ajutorul transformării:

$$\sigma_X^2=\sigma_X\sum_{i=1}^{n}\sigma_i\rho(Y_j,X) \qquad (8.4.2.1)$$

unde, pentru fiecare item i, $i=1,2,...,n$:

σ_i reprezintă abaterea standard a scorurilor observate la itemul i;

$\rho(Y_i,X)$ este coeficientul de corelaţie liniară între scorurile itemului i şi scorurile observate la întregul test.

Termenul

$$Rel(Y_i)=\sigma_i\rho(Y_i,X) \qquad (8.4.2.2)$$

se numeşte **indice de fidelitate al itemului *i*.** Valoarea sa este cu atât mai mare, cu cât este mai mare coeficientul de corelaţie liniară dintre item şi test.

Scriind formula de calcul a coeficientului α al lui Cronbach astfel:

$$\alpha=\frac{n}{n-1}\left(1-\frac{\sum_{i=1}^{n}\sigma_i^2}{\sigma_X\sum_{i=1}^{n}Rel(Y_i)}\right) \qquad (8.4.2.3)$$

se deduce că pentru a obţine un test cu o consistenţă internă bună trebuie aleşi itemi care au dispersii mici ale scorurilor şi corelaţii mari cu scorul total al testului.

Dacă un test este compus din n itemi paraleli, atunci coeficientul său de fidelitate poate fi calculat cu formula lui Spearman-Brown, 8.3.12:

$$\rho^2_{XT'} = \frac{n\rho(Y_1,Y_2)}{1+(n-1)\rho(Y_1,Y_2)}$$

din care se deduce că fidelitatea testului depinde crescător de numărul itemilor componenţi. Adăugarea de itemi paraleli va avea drept consecinţă o mărire a consistenţei interne a testului.

Din formula lui Spearman-Brown se poate afla de câte ori trebuie lungit (sau scurtat), prin itemi paraleli, un test dat, astfel încât fidelitatea testului obţinut să fie egală cu o valoare fixată. Raportul dintre lungimea noului test şi lungimea testului original se calculează cu formula:

$$k = \frac{\rho_k(1-\rho)}{\rho(1-\rho_k)} \qquad (8.4.2.4)$$

unde:
ρ = fidelitatea testului existent;
ρ_k = fidelitatea dorită, egală cu fidelitatea testului a cărui lungime este egală cu de k ori lungimea testului existent.

Practic este însă dificil să se construiască itemi de test buni, care să fie paraleli cu itemii existenţi în test. Pe de altă parte, testele lungi consumă mult timp şi sunt mai scumpe decât cele scurte. Deci, nu întotdeauna este posibilă şi rentabilă creşterea fidelităţii prin adăugarea de itemi paraleli.

8.4.3. Metoda formelor paralele

Metoda formelor paralele constă în administrarea a două teste, care reprezintă instrumente de măsură paralele, la momente foarte apropiate unul de altul şi determinarea coeficientului de corelaţie liniară între scorurile observate ale celor două forme. Valoarea calculată se numeşte **coeficient de echivalenţă**. Dacă cele două teste sunt paralele şi dacă pentru fiecare persoană erorile de

măsură la cele două administrări sunt variabile aleatoare independente, atunci coeficientul de echivalenţă coincide cu coeficientul de fidelitate al fiecărui test.

Metoda formelor paralele este potrivită în cazul când se urmăreşte măsurarea unor caracteristici *generale* ale persoanelor. Asemenea caracteristici vor influenţa în acelaşi mod scorurile ambelor teste.

Temporar	Durabil	
xxxxxxxx	xxxxxxxx	General
		Specific

Metoda formelor paralele prezintă unele avantaje faţă de metoda test-retest. Astfel, cele două forme fiind diferite, rezultatele la cel de-al doilea test nu mai sunt influenţate nici de memorarea răspunsurilor date la primul test şi nici de aflarea răspunsurilor la întrebările la care nu s-a ştiut răspunde (în cazul testelor de aptitudini sau de cunoştinţe).

Dar, metoda formelor paralele are şi inconveniente. Dintre acestea amintim următoarele:

- sunt necesare două administrări separate, ceea ce consumă timp;
- intervine oboseala sau plictiseala persoanelor examinate, ceea ce influenţează scorurile la cel de-al doilea test;
- elaborarea mai multor forme paralele de test poate fi dificilă şi costisitoare.

Un procedeu de testare asemănător cu cel descris mai sus, dar mai puţin utilizat, îl constituie administrarea celor două forme ale testului, aceluiaşi grup de persoane, la momente mai puţin apropiate. O asemenea metodă este utilă atunci când se măsoară caracteristici *generale şi durabile* ale persoanelor.

Temporar	Durabil	
	xxxxxxxx	General
		Specific

În acest caz, coeficientul de corelaţie liniară al scorurilor observate ale testelor se numeşte **coeficient de echivalenţă şi stabilitate**, întrucât el reflectă atât stabilitatea în timp a scorurilor reale ale persoanelor, cât şi echivalenţa măsurătorilor.

Scorurile obţinute la cel de-al doilea test pot să difere de cele realizate la prima testare şi datorită schimbării condiţiilor de administrare a testelor. Pentru a micşora efectele pe care această modificare le poate avea asupra

mărimii coeficientului de corelaţie dintre scorurile testelor, se obişnuieşte să se procedeze astfel (Aiken, 1994):

- Se împarte lotul de subiecţi, aleator, în două părţi egale.
- La prima testare, fiecare jumătate a lotului de subiecţi primeşte un alt test.
- La a doua testare, se inversează testele administrate, astfel încât fiecare subiect să răspundă la ambele teste.
- Se calculează coeficientul de corelaţie liniară între scorurile celor două teste, fără a se lua în considerare momentul când au fost administrate testele.

Metoda formelor paralele se foloseşte şi atunci când se doreşte obţinerea unei colecţii de teste paralele, de bună calitate. Este, de exemplu, situaţia întâlnită în psihoterapie, când este necesară administrarea periodică, la intervale de timp relativ scurte, a unor teste de anxietate sau de depresie. Pentru ca subiecţii să nu fie tentaţi să repete răspunsurile date la administrările anterioare ale testului se preferă utilizarea alternativă a unor forme de test paralele.

Un caz particular al metodei formelor paralele îl constituie **metoda înjumătăţirii** (split-half). În literatura de specialitate, aceasta este inclusă uneori în cadrul analizei consistenţei interne.

Metoda înjumătăţirii constă în următoarele:

- Se administrează testul unui lot de persoane.
- Se împarte testul în două părţi cât mai asemănătoare între ele.
- Se calculează coeficientul de corelaţie între scorurile observate la cele două jumătăţi.

Dacă jumătăţile testului sunt instrumente de măsură paralele, atunci valoarea calculată reprezintă coeficientul de fidelitate al fiecăreia dintre părţi. Cu ajutorul lui, aplicând formula lui Spearman-Brown, se poate determina coeficientul de fidelitate al întregului test.

Metoda înjumătăţirii rezolvă multe din problemele teoretice şi practice inerente metodei test-retest şi metodei formelor paralele. Mai întâi, nu este necesar să se administreze două teste diferite şi nici să se administreze de două ori acelaşi test. În felul acesta nu există pericolul modificării scorurilor reale între administrări şi nici cel al memorării şi repetării răspunsurilor date la primul test.

Dar, dificultatea constă în construirea jumătăţilor care să fie măsurători paralele. Se recomandă (Cronbach, 1943; Pitariu, 1978; Traub, 1994) să se elaboreze perechi de itemi care să fie similari ca:

- formă (mod de prezentare a textului, a ilustraţiilor);
- conţinut;
- tip de întrebare (cu răspuns liber, cu răspuns la alegere etc.);

• medie şi abatere standard a scorurilor.

Dar, itemii trebuie să fie independenţi, adică unul să nu sugereze răspunsul la celălalt.

Din fiecare pereche, un item va intra într-o jumătate de test, iar al doilea, în cealaltă jumătate.

Există multe posibilităţi de ordonare a itemilor în cadrul testului. Nu se recomandă plasarea celor două părţi una după cealaltă, întrucât intervine oboseala persoanelor examinate, care va influenţa scorurile la cea de-a doua jumătate, iar dacă testul este administrat cu limită de timp, este foarte probabil ca mulţi itemi din cea de-a doua parte să nu fie rezolvaţi datorită timpului de lucru insuficient şi nu din cauza necunoaşterii răspunsului corect. Pentru a evita asemenea situaţii, se poate adopta metoda par-impar: itemii unei jumătăţi vor ocupa locurile cu număr de ordine par, iar ceilalţi, locurile cu număr de ordine impar. Se obişnuieşte să se plaseze itemii în ordinea crescătoare a dificultăţii lor (Lord, Novick, 1968).

Apare, însă, o problemă în cazul când testul se administrează cu limită de timp. Fiecare persoană care nu reuşeşte să răspundă la ultimii $2r$ itemi va avea scorul zero la aceştia şi, deci, scorurile la itemii pari şi la cei impari vor apărea mai puternic corelate decât sunt în realitate. Din acest motiv, metoda înjumătăţirii nu se va aplica atunci când administrarea testului se face cu limită de timp.

8.4.4. Coeficienţi de fidelitate inter-evaluatori

În cazul testelor care nu au o cotare obiectivă, cum sunt testele proiective sau cele de creativitate, scorurile subiecţilor sunt influenţate şi de persoana care face evaluarea răspunsurilor. Chiar dacă evaluatorii îndeplinesc cerinţele formulate în manualul testului, referitoare la cunoştinţele şi experienţa necesare, este puţin probabil ca ei să atribuie scoruri identice fiecărui subiect.

Pentru a verifica în ce măsură scorurile testului sunt dependente de cel care a făcut cotarea, se calculează coeficientul de corelaţie liniară între scorurile atribuite aceloraşi subiecţi de evaluatori diferiţi. Un asemenea coeficient se numeşte **coeficient de fidelitate inter-evaluatori**.

Unii autori (Hammond, 1995) recomandă utilizarea, pentru aprecierea acordului dintre evaluatori, a coeficientului K al lui Cohen. Acesta ia valori între -1 şi 1 şi indică în ce măsură concordanţa evaluărilor este diferită de cea datorată întâmplării (A se vedea Anexa I).

8.5. OBSERVAŢII ASUPRA METODELOR FOLOSITE PENTRU ESTIMAREA FIDELITĂŢII

Alegerea metodei aplicate pentru estimarea fidelităţii trebuie să ţină seama de ceea ce măsoară testul, de scopul pentru care el va fi folosit şi de condiţiile concrete în care se poate face testarea (există sau nu posibilitatea administrării repetate a testului, există sau nu forme paralele ale testului etc.).

Atunci când testul măsoară o caracteristică durabilă şi, deci, trebuie să fie puţin sensibil la situaţia de testare, sau când pe baza scorurilor testului urmează să se tragă anumite concluzii valabile pentru o perioadă de timp mai îndelungată, se va aplica metoda test-retest.

Dacă nu se poate asigura retestarea persoanelor după un interval de timp suficient de lung pentru ca subiecţii să nu îşi mai poată aminti răspunsurile date la test, dar există o formă paralelă a testului, se va utiliza metoda formelor paralele.

În cazul particular când forma paralelă a testului a fost administrată împreună cu testul se va putea aplica metoda înjumătăţirii.

Dacă se ştie că testul se compune din itemi care măsoară aceeaşi caracteristică -mai mult sau mai puţin durabilă- pentru estimarea coeficientului de fidelitate se va folosi o formulă de calcul al unui coeficient de consistenţă internă.

În cazul testelor pentru care cotarea nu este obiectivă, evaluatorul devine o componentă a procesului de măsurare realizat cu ajutorul testului. Alături de o altă meodă de estimare a fidelităţii se va calcula şi coeficientul de fidelitate inter-evaluatori.

În general valorile coeficienţilor de consistenţă internă sunt mai mari decât cele ale coeficienţilor de echivalenţă, care sunt, de obicei, mai mari decât coeficienţii de stabilitate (Murphy, Davidshofer, 1987). Acest lucru se explică prin faptul că numărul factorilor care influenţează erorile de măsură este mai mare în metoda test-retest şi în metoda formelor paralele, decât în metoda analizei consistenţei interne. De exemplu, acumularea de informaţii între test şi retest va afecta numai valoarea coeficienţilor de stabilitate, nu şi pe aceea a coeficienţilor de echivalenţă sau a coeficienţilor consistenţei interne.

Dacă, însă, persoanele examinate ţin minte răspunsurile date la test şi le repetă la retest, se poate obţine un coeficient de stabilitate mare, chiar dacă testul are consistenţă internă scăzută.

8.6. FACTORI CARE INTERVIN ÎN ESTIMAREA COEFICIENŢILOR DE FIDELITATE

Trebuie observat că experimentele efectuate pentru determinarea valorii unui coeficient de fidelitate nu cuprind, în general, întreaga populaţie de persoane căreia îi este destinat testul, ci doar un eşantion din această populaţie. Prin urmare, valoarea calculată este doar o aproximare (o estimare) a coeficientului de fidelitate. Luând un alt eşantion, se va obţine o altă valoare.

Pentru ca estimarea coeficientului de fidelitate să fie cât mai precisă, este necesar să fie îndeplinite următoarele condiţii (Traub, 1994):

A. Eşantionul să fie cât mai voluminos. Deşi forma exactă a repartiţiei coeficienţilor de fidelitate calculaţi pentru diverse eşantione de acelaşi volum, extrase aleator din populaţie, nu este cunoscută, se obişnuieşte să se considere că abaterea standard a acestei repartiţii (denumită eroare standard) este invers proporţională cu rădăcina pătrată din volumul eşantioanelor. Prin urmare, pentru a micşora de două ori eroarea standard este necesar să se mărească de patru ori volumul eşantionului.

B. Eşantionul de subiecţi să fie reprezentativ pentru populaţia căreia îi este destinat testul. El trebuie să fie la fel de eterogen ca şi această populaţie, întrucât gradul de omogenitate al eşantionului are influenţă asupra mărimii coeficientului de fidelitate.

C. Măsurătorile efectuate cu ajutorul testului să fie independente între ele, adică măsurătoarea realizată asupra unui subiect să nu fie influenţată şi nici să nu influenţeze măsurătorile efectuate asupra altor persoane. Dacă administrarea testului se face în grup, acest lucru înseamnă că persoanele examinate vor fi supravegheate astfel încât nici una să nu poată copia.

Atunci când calculul coeficientului de fidelitate impune ca fiecare persoană să răspundă de două ori la acelaşi test sau să răspundă la două forme paralele ale testului foarte asemănătoare între ele, cerinţa de independenţă a măsurătorilor se extinde, în sensul că măsurătorile efectuate asupra aceleiaşi persoane trebuie să fie independente între ele. Aceasta înseamnă că nu li se va oferi subiecţilor posibilitatea de a revedea sau de a-şi aminti răspunsurile date la prima administrare a testului.

Dacă se calculează un coeficient de consistenţă internă, se va urmări ca itemii testului să fie independenţi între ei. Prin urmare, rezolvarea unuia nu va fi condiţionată de rezolvarea altuia şi nu vor exista itemi care să se refere la

acelaşi aspect, astfel încât subiecţii să răspundă la ei în acelaşi mod (corect sau greşit).

D. Procedura de administrare a testului şi a retestului sau a celor două forme paralele ale testului să fie identică şi să fie aceeaşi cu cea care se va utiliza în practică. Astfel, de exemplu, dacă se prevede ca testul să fie administrat cu limită de timp, în 40 de minute, atunci acest interval de timp va fi respectat şi la administrarea testului pentru estimarea fidelităţii.

8.7. INTERPRETAREA COEFICIENŢILOR DE FIDELITATE

Dintr-o mulţime de teste având toate celelalte caracteristici egale, este de preferat testul care are cea mai mare fidelitate. Dar, pe de o parte, testul cel mai fidel poate fi cel mai scump sau mai dificil de administrat. Pe de altă parte, caracteristicile testelor sunt rareori egale.

Este imposibil să se specifice un nivel minim care să fie impus coeficienţilor de fidelitate ai tuturor testelor. Dar, în practica obişnuită sunt utile următoarele recomandări:

A. Se vor cere niveluri înalte de fidelitate testelor folosite pentru a se lua decizii importante asupra persoanelor şi celor care împart indivizii în mai multe categorii pe baza unor diferenţe relativ mici între ei. În selecţia profesională, de pildă, se vor folosi teste care au coeficienţi de fidelitate mai mari decât 0.9 (J. C. Nunnally, după Hammond, 1995).

B. Se vor accepta niveluri scăzute de fidelitate atunci când testele se utilizează pentru a lua decizii preliminare, nu finale, şi în cazul testelor care împart persoanele într-un număr mic de categorii, pe baza unor diferenţe individuale mari.

C. Dacă un test serveşte la compararea grupelor de persoane, coeficientul său de fidelitate poate fi mai mic, chiar cuprins între 0.6 şi 0.7. Dacă, însă testul se foloseşte pentru compararea persoanelor, una cu alta, coeficientul său de fidelitate trebuie să fie mai mare decât 0.85 (Aiken, 1994).

În general, fidelitatea chestionarelor de personalitate şi a celor de interese este mai mică decât cea a testelor cognitive (de cunoştinţe, de inteligenţă, de aptitudini speciale) (Aiken, 1994).

În practică se întâlnesc de obicei coeficienţi de fidelitate cuprinşi între 0.70 şi 0.98 (Guilford, 1965). Testele cotate obiectiv care măsoară aptitudini cognitive pot furniza, în loturi eterogene de persoane, coeficienţi de fidelitate mai mari decât 0.8. Testele cotate subiectiv care măsoară aptitudini, testele de cunoştinţe şi cele care evaluează performanţe tipice (de exemplu chestionare de personalitate) au, adesea, coeficienţi de fidelitate mai mici decât 0.80 (Traub, 1994).

Testele standardizate de cunoştinţe au, în general, o fidelitate mare spre moderată. În schimb, testele cu alegere multiplă utilizate la clasă pot avea o fidelitate mai scăzută. Coeficientul lor de fidelitate este apropiat de 0.75.

Validitatea fiind mai importantă decât fidelitatea, în practică se folosesc adesea teste care sunt valide pentru populaţia şi situaţia în care sunt administrate, dar care nu au coeficienţi de fidelitate foarte mari.

8.8. FIDELITATEA DIFERENŢEI SCORURILOR ÎNTRE TEST ŞI RETEST

De multe ori ne interesează modul în care variază scorurile observate la un test $m \in \mathbf{M}$, administrat de două ori unui lot de persoane **P** (de exemplu, înainte şi după un tratament). Procedeul de lucru obişnuit constă în calculul mediei diferenţelor individuale dintre scorul retestului şi cel al testului şi aplicarea unui test statistic pentru verificarea semnificaţiei acestei medii. Se pune, însă, problema: cât de bine măsoară diferenţele de scor schimbările reale petrecute în indivizi? Altfel spus, cât de fidelă este măsurarea diferenţelor de scor?

Pentru a studia fidelitatea diferenţelor de scor se defineşte, pentru fiecare persoană $p \in \mathbf{P}$, variabila aleatoare D_{mp}, ale cărei valori sunt toate diferenţele de scor care ar putea fi observate la acea persoană:

$$D_{mp} = X_{mp,2} - X_{mp,1} \qquad (8.8.1)$$

$X_{mp,1}$ şi $X_{mp,2}$ sunt variabilele aleatoare corespunzătoare scorurilor observate la test, respectiv la retest.

Făcând presupunerea obişnuită din teoria testelor, variabila D_{mp} se compune dintr-o "valoare reală a diferenţei scorurilor", T_{mp} şi o "eroare de măsură a diferenţei scorurilor", E_{mp}:

$$D_{mp} = T_{mp} + E_{mp} \qquad (8.8.2)$$

Variabila T_{mp} este definită ca şi medie a diferenţelor observate:

$$T_{mp} = \mathbf{M}(D_{mp}) \qquad (8.8.3)$$

deci ea este egală cu diferenţa scorurilor reale ale persoanei p între test şi retest.

Se notează:
D_{m*} = variabila aleatoare corespunzătoare diferenţei scorurilor observate în întreaga populaţie **P**;
T_{m*} = variabila aleatoare care ia ca valori diferenţele scorurilor reale între test şi retest în populaţia **P**;
E_{m*} = variabila eroare de măsură pentru diferenţele de scor în populaţia **P**.

Conform teoremei 6.1.1.1 în populaţia **P** are loc relaţia

$$D_{m*} = T_{m*} + E_{m*} \tag{8.8.4}$$

şi se poate calcula fidelitatea diferenţelor de scor pe baza formulei de definiţie a coeficientului de fidelitate:

$$\rho^2(D_{m*}, T_{m*}) = \frac{\sigma^2(T_{m*})}{\sigma^2(D_{m*})} \tag{8.8.5}$$

Ţinând seama de felul în care a fost definită variabila D_{mp}, se poate rescrie formula (8.8.5) astfel încât fidelitatea diferenţelor de scor să fie exprimată cu ajutorul fidelităţii testului, ρ_1^2, al fidelităţii retestului, ρ_2^2, al dispersiilor în populaţia **P** a scorurilor observate la test şi la retest, σ^2_1 şi σ^2_2, precum şi a covarianţei scorurilor observate la cele două administrări ale testului, σ_{12} (Traub, 1994):

$$\rho^2(D_{m*}, T_{m*}) = \frac{\sigma_1^2\rho_1^2 + \sigma_2^2\rho_2^2 - 2\sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}} \tag{8.8.6}$$

Se constată de aici că fidelitatea diferenţelor de scor depinde crescător de fidelitatea testului şi de cea a retestului, şi depinde descrescător de mărimea covarianţei scorurilor observate la test şi la retest. Cu cât fidelitatea testului şi cea a retestului sunt mai bune, cu atât fidelitatea diferenţelor de scor este mai bună. Când scorurile observate la test şi la retest sunt necorelate liniar ($\sigma_{12}=0$), fidelitatea diferenţelor de scor este maximă.

CAPITOLUL 9
VALIDITATEA TESTELOR PSIHOLOGICE

9.1. DEFINIŢIA VALIDITĂŢII

Validitatea este unul dintre conceptele cele mai importante în activitatea de cercetare ştiinţifică. Ori de câte ori un cercetător îşi pune o întrebare de forma "Acest studiu răspunde într-adevăr problemei formulate ?" sau "Acest instrument măsoară cu adevărat ceea ce vrem să măsurăm ?" sau "Cercetările din acest laborator ne arată într-adevăr lumea reală?", el se interesează de fapt de validitatea studiului ori a instrumentului.

În esenţă, **validitatea** se referă la cât de bine realizează un studiu, o procedură sau un instrument de măsură ceea ce se presupune că face (Graziano, Raulin, 1993). În acest sens a fost înţeleasă validitatea, multă vreme, şi în teoria testelor. Iniţial, ea a fost definită şi cercetată în legătură cu funcţia îndeplinită de test. Astfel, de exemplu, Lafon (1973) definea validitatea unui test ca fiind calitatea acestuia "de a măsura exact ceea ce trebuie să măsoare", iar Lindeman (1978) scria că "validitatea unui test este dată de măsura în care acesta îşi îndeplineşte funcţia".

Observaţia că psihologii care utilizează un test nu sunt interesaţi, de fapt, de proprietăţile acestuia, ci de atributele persoanelor testate, a condus spre o nouă accepţiune a termenului "validitate". Acum accentul nu mai cade pe integritatea testului, ci pe încrederea pe care o putem avea în deducţiile făcute pornind de la scoruri, luând în considerare întregul proces prin care s-au obţinut aceste scoruri (Guion, 1986). Prin urmare, validitatea nu mai este privită ca o calitate a testului, ci ca o calitate a utilizării date rezultatelor testului. Sunt relevante, în acest sens, următoarele definiţii:

- Validitatea este gradul în care o anumită inferenţă făcută plecând de la test are înţeles, este potrivită şi utilă (*Standards for Educational and Psychological Tests*, 1985).
- Validitatea este o judecată evaluativă globală asupra gradului în care dovezi empirice şi raţionamente teoretice oferă sprijin ca, pe baza scorurilor testului, să se realizeze interpretări sau acţiuni potrivite (Messick, 1995).
- Validitatea unui test se referă la cât de potrivite sunt interpretările descriptive, explicative sau predictive care se dau scorurilor sale (Silva, 1993).

În această nouă accepţiune, validitatea unui test trebuie legată atât de test, de ceea ce măsoară itemii săi, cât şi de contextul în care este utilizat testul şi de persoanele care răspund la test. Este absolut necesar ca atunci când se

vorbeşte despre validitatea unui test să se precizeze funcţia la care se referă şi atributul măsurat sau variabila criteriu utilizată la compararea scorurilor. Nu este corect să se folosească expresia "validitatea testului", fără alte specificaţii. Nici un test nu este valid pentru toate scopurile sau în toate situaţiile sau pentru toate eşantioanele de indivizi (*Standards for Educational and Psychological Tests*, 1974).

Exemplu.

Să considerăm un test de cunoştinţe de matematică, ai cărui itemi conţin exerciţii cu integrale.

Pentru a putea aprecia validitatea acestuia este necesar, în primul rând, să se analizeze testul, urmărind dacă măsurarea se face corect, adică dacă răspunsurile notate în grila testului sunt corecte şi dacă itemii măsoară ceea ce se doreşte. De exemplu, dacă se intenţionează verificarea cunoaşterii formulelor de calcul, itemii trebuie să conţină exerciţii care se rezolvă prin simpla aplicare a acestor formule. În schimb, dacă se urmăreşte testarea deprinderii de a opera cu integrale, itemii vor conţine exerciţii care necesită efectuarea unor transformări pentru a se ajunge la formulele de calcul învăţate.

În continuare se ia în considerare contextul în care se va aplica testul. El poate servi, de pildă, la evaluarea cunoştinţelor de matematică ale absolvenţilor de liceu, la examenul de bacalaureat. Dar dacă se doreşte să fie administrat la concursul de admitere la facultatea de psihologie, pentru a selecta persoanele care vor putea utiliza corect procedee statistice în prelucrarea datelor provenite din experimentele psihologice, este posibil ca predicţiile făcute pe baza scorurilor sale să nu fie corecte. Ar fi mai indicat să se folosească, în acest scop, un test de cunoştinţe care să conţină exerciţii cu probabilităţi.

Pe de altă parte, un test care conţine exerciţii cu integrale poate fi util pentru verificarea cunoştinţelor de matematică ale candidaţilor la Universitatea Tehnică. Dar, dacă exerciţiile sunt de dificultate medie, atunci la facultăţile cu concurenţă foarte mare, unde se înscriu absolvenţi de liceu foarte bine pregătiţi, şi la cele cu concurenţă foarte mică, unde se înscriu candidaţi cu o pregătire slabă, scorurile obţinute de subiecţi vor fi apropiate ca valoare şi nu vor permite o diferenţiere între persoane.

Dacă testul este administrat cu ajutorul calculatorului, atunci rezultatele sale pot fi influenţate şi de factori care nu au legătură cu cunoştinţele de matematică (de pildă, de anxietatea faţă de calculator, care este legată de nivelul cunoştinţelor de informatică, sau de deprinderea subiecţilor de a folosi tastatura). Prin urmare, este greşit să se aprecieze nivelul cunoştinţelor de matematică pe baza scorurilor testului, dacă subiecţii diferă între ei în privinţa experienţei în lucrul cu calculatorul.

Validarea unui test este procesul prin care se investighează gradul de validitate a interpretării propuse de test (*Standards for Educational and Psychological Tests*, 1985). Orice validare trebuie să înceapă prin a răspuunde la întrebările:

- Ce funcţie îndeplineşte testul?
- În ce scop se utilizează testul?
- În ce condiţii se foloseşte testul?
- Ce caracteristici au persoanele cărora li se administrează testul?

Multă vreme, procedeele aplicate pentru validarea testelor au fost legate de răspunsul dat primei întrebări. Pentru fiecare dintre funcţiile testelor psihologice - măsurare, decizie şi predicţie - au fost formulate condiţiile pe care trebuie să le îndeplinească un test valid şi au fost propuse strategii de verificare a acestora. Fiecare set de condiţii, cu strategiile corespunzătoare, a fost etichetat şi a fost privit ca un "tip" de validitate sau ca o "faţă" a validităţii. În *Standards for Educational and Psychological Tests* din anul 1974 sunt prezentate trei **tipuri de validitate**:

- validitatea conceptuală (construct validity);
- validitatea de conţinut (content validity);
- validitatea relativă la criteriu (criterion-oriented validity).

Validitatea conceptuală şi validitatea de conţinut sunt legate de funcţia de măsurare a testelor. Un test este valid ca instrument de măsură dacă prin scorurile sale se realizează o măsurare corectă a atributului psihic pentru care a fost construit. Aprecierea corectitudinii măsurării impune identificarea tipului de scală pe care se face măsurarea şi verificarea condiţiilor din definiţia acestuia.

Validitatea relativă la criteriu este legată de funcţia de decizie şi de cea de predicţie. Un test este valid relativ la criteriu dacă pe baza scorurilor sale se pot lua decizii corecte sau se pot face prognoze corecte asupra persoanelor examinate. Strategia utilizată pentru a aprecia dacă un test este valid relativ la criteriu constă în compararea scorurilor sale cu valorile unei variabile criteriu. Se apreciază că testul este valid dacă el poate înlocui variabila criteriu, adică dacă asemănarea dintre scorurile sale şi valorile variabilei criteriu este atât de mare încât, cunoscând scorul la test al unei persoane, se poate deduce valoarea variabilei criteriu pentru aceasta.

În practică nu este posibil totdeauna să se obţină valorile variabilei criteriu concomitent cu cele ale testului. Din acest motiv, strategiile folosite pentru aprecierea validităţii relative la criteriu au fost împărţite în două clase. Au apărut, astfel, două tipuri noi de validitate care, în unele lucrări de specialitate, înlocuiesc validitatea relativă la criteriu. Este vorba despre:

- **validitatea concurentă**, care necesită obţinerea scorurilor la criteriu aproximativ în acelaşi timp cu cele ale testului;
- **validitatea predictivă**, care permite ca administrarea testului să preceadă înregistrarea valorilor variabilei criteriu, între cele două acţiuni existând un interval de timp mai lung sau mai scurt, în funcţie de situaţia concretă.

Standards for Educational and Psychological Tests din 1985 afirmă că nu se poate vorbi despre "tipuri" diferite de validitate: cele trei (sau patru) tipuri descrise anterior sunt doar simple etichete atribuite unor categorii de strategii de validare care se pot utiliza şi asociate. De pildă, pentru un test de atenţie concentrată se pot cerceta atât validitatea conceptuală (constructul măsurat fiind atenţia concentrată), cât şi validitatea relativă la criteriu (drept criteriu poate servi testul de atenţie concentrată Toulouse-Pieron).

Ideea asocierii fiecărei categorii de strategii de validare cu o anumită funcţie a testelor a fost abandonată. Se acceptă acum că validarea relativă la criteriu poate fi efectuată atât în cazul testelor care măsoară un construct, cât şi pentru testele utilizate în predicţie (Guion, 1986). În primul caz, criteriul este un construct ales datorită legăturilor pe care se presupune că le are (sau că nu le are) cu constructul măsurat de test. În cel de-al doilea caz, criteriul este o măsură a performanţei sau a comportamentului în muncă. Nu se poate afirma că o strategie de validare este mai potrivită decât alta pentru o anumită utilizare a testelor sau pentru o categorie specifică de inferenţe bazate pe scoruri.

C. L. Lawshe se pronunţă pentru înlocuirea denumirii "tip de validitate" cu cea de "tip de analiză a validităţii"; prin "analiză a validităţii" el înţelege o procedură, un proces sau o strategie prin care se colectează sau se generează date în vederea determinării extinderii ori a forţei inferenţelor care pot fi făcute pe baza unei mulţimi de scoruri de test (Landy, 1986). Cu o singură modificare minoră, ce constă în înlocuirea expresiei "mulţime de scoruri de test" cu aceea de "mulţime de observaţii", definiţia dată de Lawshe se transformă într-o definiţie a testării ipotezelor. Se constată, astfel, că validarea testelor este o operaţie similară verificării ipotezelor ştiinţifice. În cazul validării testelor, ipotezele pot fi formulate astfel:

> "Persoanele care au realizat un scor mare la testul X vor realiza o performanţă bună în activitatea Y."

sau

> "Persoanele care au obţinut scoruri mari la testul X au un nivel înalt al constructului Y."

În continuare vom prezenta categoriile de strategii de validare. În locul denumirilor "tipurilor" de validitate întâlnite în majoritatea lucrărilor de specialitate din limba română:

- "validitate conceptuală" (concept validity) sau "validitate de construct" (construct validity)

şi

- "validitate de conţinut" (content validity)

vom utiliza denumirile

- "validitate relativă la construct"

şi

- "validitate relativă la conţinut",

care, alături de denumirea consacrată "validitate relativă la criteriu", sugerează că validitatea testului este cercetată sub unul dintre aspectele posibile (construct, conţinut sau criteriu).

9.2. VALIDITATEA TESTULUI RELATIVĂ LA CONSTRUCTUL MĂSURAT DE ACESTA

Validitatea relativă la construct se aplică în cazul testelor care măsoară constructe psihologice, cum sunt chestionarele de personalitate. Ea face apel la diverse teorii psihologice pentru a verifica, în primul rând, dacă testul se referă cu adevărat la constructul pe care trebuie să îl măsoare, şi apoi, dacă scorurile obţinute de subiecţi la test reflectă corect mărimea acestui construct la persoanele respective; altfel spus, dacă între scorurile testului şi valorile constructului există o relaţie funcţională monotonă (de obicei această relaţie este presupusă liniară):

$$\text{scor la test} = f(\text{valoarea constructului})$$

Validitatea relativă la construct impune analizarea testului din mai multe puncte de vedere şi prin metode diferite. Un algoritm posibil al validării relative la construct, valabil pentru toate măsurătorile psihologice şi educaţionale, inclusiv pentru evaluările performanţei, constă din următoarii paşi (după Messick, 1995):

A. Mai întâi se cercetează cât de relevant şi reprezentativ este conţinutul testului pentru constructul care trebuie măsurat. O asemenea analiză impune descrierea amănunţită a *domeniului constructului* respectiv, adică prezentarea deprinderilor, a cunoştinţelor, atitudinilor, motivelor, a altor atribute care au legătură cu constructul şi specificarea relaţiilor dintre ele. Un grup de experţi apreciază apoi

dacă cele mai importante porţiuni ale domeniului constructului sunt acoperite de conţinutul testului.

Este foarte utilă şi cercetarea modului în care constructul este "reprezentat în test". Pentru aceasta se realizează un model al rezolvării testului, conţinând mecanismele teoretice (procese, strategii, cunoştinţe) despre care se presupune că ar avea influenţă asupra răspunsurilor date la test. Se analizează apoi dacă elementele modelului au legătură cu constructul pe care trebuie să îl măsoare testul şi cu răspunsurile care se pot obţine la test. De mare folos, în această analiză, poate fi reţeaua nomologică întocmită cu ocazia explicării constructului.

Dacă scorurile testului sunt dependente mai mult de alţi factori decât de componentele constructului care trebuie măsurat, se vorbeşte despre o "subreprezentare a constructului în test". Acest lucru se poate întâmpla dacă elemente care nu au legătură cu constructul determină ca rezolvarea testului să fie foarte dificilă sau, din contră, foarte uşoară. De exemplu, dacă itemii unui test de cunoştinţe de statistică pretind efectuarea unor calcule cu numere având multe zecimale, apare riscul de a obţine răspunsuri eronate din cauza greşelilor de calcul şi nu datorită absenţei cunoştinţelor de statistică.

B. O altă etapă a validării constă în analizarea consecvenţei răspunsurilor la itemii testului. În acest scop se identifică procesele implicate în obţinerea scorurilor la itemi şi se urmăreşte dacă acestea acţionează similar la toţi itemii din întregul test sau dintr-o parte fixată a acestuia.

Mijloacele de verificare a consecvenţei răspunsurilor sunt variate: se cercetează dacă timpii de răspuns la itemi sunt aproximativ constanţi sau dacă scorurile itemilor corelează semnificativ între ele ori dacă mişcările ochilor subiecţilor în timpul conceperii răspunsurilor nu variază prea mult de la un item la altul.

C. Pornind de la ideea că scorul obţinut la test este rezultatul acţiunii mai multor variabile, dintre care unele au legătură cu constructul măsurat, se urmăreşte dacă structura scorului la test reflectă structura domeniului constructului măsurat.

Pentru aceasta, se descompune scorul total al testului în scoruri parţiale (scoruri ale unor grupe de itemi care corelează între ei sau scoruri factoriale) şi se presupune că fiecare asemenea scor reprezintă măsura unei variabile.

Adesea, în scopul identificării unor asemenea variabile se utilizează analiza factorială. Ea permite să se determine una sau mai multe variabile (numite factori) ale căror valori (numite scoruri factoriale) se combină liniar pentru a forma scorul total al testului. Această metodă, numită *validare factorială*, se aplică mai ales în cazul testelor care măsoară constructe multidi-

mensionale, când se verifică şi dacă variabilele puse în evidenţă de analiza factorială corespund componentelor constructului.

Se caută semnificaţia variabilelor măsurate prin scorurile parţiale ale testului, se reţin cele care au legătură cu constructul şi se verifică dacă importanţa acestora în domeniul constructului este reflectată corect prin ponderile pe care le au măsurile lor în scorul total al testului.

D. Se estimează în ce măsură concluziile desprinse până la acest pas al algoritmului de validare pot fi generalizate, adică pot fi extinse asupra altor populaţii sau altor ocazii de examinare cu ajutorul testului. Se încearcă să se stabilească dacă interpretările propuse pentru scorurile testului se menţin atunci când se schimbă situaţia în care se face examinarea, momentul examinării ori evaluatorii.

E. Se cercetează relaţiile dintre scorurile testului şi diverse alte măsurători psihologice sau observaţii non-evaluative asupra comportamentului. Această verificare este denumită uneori *validare convergentă şi discriminantă*.

Un test are *validitate convergentă* dacă măsoară ceea ce evaluează şi alte teste sau variabile care se referă la acelaşi construct, adică între scorurile sale şi rezultatele respectivelor măsurători există o relaţie funcţională monotonă (liniară sau nu). Un test are *validitate discriminantă* dacă evaluează altceva decât diverse teste sau variabile despre care se ştie că se referă la constructe ce nu au legătură cu constructul măsurat de test. Aceasta înseamnă că scorurile testului nu sunt în relaţie funcţională monotonă cu rezultatele acelor măsurători.

Verificarea validităţii convergente şi discriminante se realizează conform următorului algoritm:

(1) Se identifică acele comportamente care sunt legate de constructul măsurat de test.

(2) Se determină constructele care au legături cu constructul măsurat de test şi se pun în evidenţă acele comportamente legate de ele care pot fi conectate cu constructul măsurat de test.

(3) Pentru toate comportamentele reţinute se apreciază ce fel de corelaţii ar putea exista între măsurile lor şi o măsură bună a constructului care ne interesează (corelaţii pozitive, negative, mari, mici sau nule).

(4) Se determină coeficienţii de corelaţie între scorurile la testul studiat şi rezultatele unor măsurători (eventual scoruri de teste) efectuate asupra comportamentelor reţinute.

(5) Se compară coeficienţii de corelaţie obţinuţi cu valorile aşteptate.

O metodă utilizată pentru verificarea validităţii convergente şi discriminante a testelor este aceea propusă de D. T. Campbell şi D. W. Fiske

(Anastasi, 1976), bazată pe matricea "trăsături multiple - metode multiple" (multitrait-multimethod matrix). Aceasta conţine valorile coeficienţilor de corelaţie liniară între diverse măsurători ale aceloraşi constructe, obţinute prin intermediul a două sau trei tehnici ori instrumente diferite.

Exemplu

În urma evaluării constructelor A, B şi C, despre care se face presupunerea că nu au legături între ele, prin trei metode diferite -"test I", "test II" şi "test III"- s-a obţinut următoarea matrice de coeficienţi de corelaţie liniară (pentru a se putea urmări mai uşor datele, ea a fost scrisă sub forma unui tabel):

		Test I A	Test I B	Test I C	Test II A	Test II B	Test II C	Test III A	Test III B	Test III C
Test I	A	.79								
	B	.10	.85							
	C	.03	.34	.75						
Test II	A	.42			.76					
	B	.12	.45		.03	.77				
	C	.06	.19	.56	.12	.17	.83			
Test III	A	.50			.54			.86		
	B	.33	.45		.39	.66		.38	.84	
	C	.25	.40	.58	.29	.37	.63	.35	.41	.85

Observaţie. Coeficienţii cu valori mai mari decât .35 sunt semnificativi la pragul p=0.05.

Valorile notate pe diagonala principală sunt coeficienţii de fidelitate ai testelor, obţinuţi prin metoda formelor paralele. De exemplu, pentru testul II, coeficienţii de fidelitate sunt 0.76 (pentru constructul A), 0.77 (pentru constructul B) şi 0.83 (pentru constructul C).

Valorile de pe diagonalele celorlalte căsuţe sunt coeficienţii de corelaţie între măsurătorile realizate de cele trei teste pentru aceleaşi constructe. Ei servesc la verificarea validităţii convergente. De exemplu, pentru constructul A, coeficienţii de corelaţie sunt:

0.42 (testul I cu testul II);
0.50 (testul I cu testul III);
0.54 (testul II cu testul III).

Toţi aceşti coeficienţi sunt semnificativi statistic la pragul p=0.05, ceea ce denotă o bună validitate convergentă a celor trei teste.

Celelalte valori din matrice reprezintă coeficienţii de corelaţie liniară între constructe diferite, măsurate prin acelaşi test sau prin teste diferite. Ei servesc la aprecierea validităţii discriminante a testelor.

Se observă că testele I şi II discriminează între constructele A, B şi C (coeficienţii de corelaţie aflaţi sub diagonala căsuţelor I-I, I-II şi II-II sunt nesemnificativi la pragul p=0.05), în timp ce testul III nu are validitate discriminantă: coeficienţii de corelaţie liniară între constructele A, B şi C, măsurate prin testul III (în căsuţa III-III) sau prin testul III şi un alt test (în căsuţele I-III şi II-III) sunt, în general, semnificativi la pragul p=0.05.

În exemplul prezentat, validitatea convergentă şi discriminantă a testelor I, II şi III a fost uşor de dedus. Dar, în practică, apar situaţii complexe, greu de interpretat. Uneori corelaţiile dintre constructe diferite măsurate prin aceeaşi metodă sunt mai mari decât corelaţiile dintre măsurile aceluiaşi construct obţinute prin metode diferite (chestionar creion-hârtie, tehnică proiectivă, scală de evaluare, interviu etc.). Aceasta înseamnă că metodele, mai precis unele elemente ale metodelor, au o importanţă mai mare în evaluare decât ceea ce se măsoară prin ele.

Mai trebuie reţinut că pentru aprecierea validităţii discriminante a unui test este util să se calculeze, pe lângă coeficientul de corelaţie dintre test şi o altă măsurătoare, şi coeficientul de corelaţie dintre constructele măsurate. Se aplică formula 8.2.2. Întrucât coeficienţii de fidelitate iau valori subunitare (valoarea 1 este extrem de rară), corelaţia constructelor este, în valoare absolută, mai mare decât valoarea absolută a coeficientului de corelaţie al scorurilor observate. Dacă se obţine un coeficient de corelaţie mare între două constructe care, teoretic, nu au nimic în comun, validitatea discriminantă a testului este necorespunzătoare.

Validarea convergentă a unui test este necesară, dar nu şi suficientă pentru a dovedi că testul măsoară ceea ce se doreşte. Aceasta, deoarece etichetele atribuite testelor reprezintă trăsăturile sau predispoziţiile pe care testele *ar trebui* să le măsoare. Corelaţia mare între două teste care poartă aceeaşi etichetă poate fi cauzată de existenţa, în cele două teste, a unor itemi asemănători, care nu totdeauna sunt reprezentativi pentru constructul pe care testele intenţionează să îl măsoare. Analiza factorială sau analiza de clusteri efectuată asupra itemilor din testele cu aceeaşi denumire poate ajuta la explicarea corelaţiilor dintre scorurile testelor.

F. Se analizează interpretările propuse pentru scorurile testului şi sursele de invaliditate a testului, pentru a deduce posibilele consecinţe negative ale utilizării testului, atât pe termen scurt, cât şi pe perioade mai îndelungate.

Din cele prezentate mai sus se constată că întregul proces de validare relativă la construct a testului constă în formularea de ipoteze ştiinţifice (asupra componentelor constructului, asupra relaţiilor dintre aceste componente etc.) şi în verificarea acestora. Metodele aplicate în validare sunt dependente de ipotezele formulate. Din acest motiv nu există metode valabile pentru validarea relativă la construct a oricărui test.

Trebuie reţinut că prin validarea testului relativă la constructul măsurat se validează atât testul cât şi ipotezele aflate la baza sa. Validitatea testului şi validitatea constructului sunt inseparabile. Când un test nou este elaborat pentru măsurarea unui construct cunoscut, riscul ca validitatea să fie nesatisfăcătoare este mai mare pentru test decât pentru construct. Dar, poate să se constate şi necesitatea revizuirii constructului (L.J. Cronbach, după Silva, 1993).

9.3. VALIDITATEA TESTULUI RELATIVĂ LA CONŢINUTUL SĂU

Acest tip de validitate este analizat la testele care se utilizează pentru a estima "cum acţionează o persoană în universul de situaţii pe care testul intenţionează să îl reprezinte" (*Standards for Educational and Psychological Tests*, 1974).

În literatura psihologică există păreri diferite despre ceea ce se urmăreşe prin validarea testului relativă la conţinut. Anastasi (1976), de pildă, consideră că acest tip de validare implică "examinarea sistematică a conţinutului testului pentru a determina dacă el acoperă un eşantion reprezentativ din domeniul de comportamente pe care testul în cauză trebuie să îl măsoare". În această accepţiune, itemii testului sunt priviţi ca reprezentând un eşantion extras dintr-o populaţie largă de itemi posibili care definesc ceea ce testul trebuie să măsoare. Dacă eşantionul este reprezentativ pentru populaţie, atunci se admite că testul are validitate relativă la conţinut (Gregory, 1992).

Alţi autori înlocuiesc expresia "domeniu de comportamente" cu "mulţimea sarcinilor" sau "mulţimea performanţelor".

R. T. Lennon consideră că validitatea relativă la conţinut se referă la răspunsurile subiectului şi nu la întrebările testului. Pentru analizarea acestui tip de validitate trebuie luat în considerare nu numai conţinutul itemilor testului ci

şi procesul utilizat de subiect pentru a ajunge la formularea răpunsului (Silva, 1993). De exemplu, un test având drept domeniu de conţinut "cunoştinţele de ortografie", format din itemi cu alegere multiplă, poate verifica dacă un individ recunoaşte cuvintele scrise corect din punct de vedere ortografic, dar nu şi dacă acesta ştie să scrie corect după dictare. Deci nu este suficient să se spună că testul evaluează cunoştinţele de ortografie.

Adoptând punctul de vedere formulat de Anastasi (1976), analiza validităţii relative la conţinut a unui test impune ca prim pas definirea şi descrierea domeniului de conţinut al testului. Este esenţial ca descrierea să fie cât mai detaliată şi să fie precizate clar limitele acestui domeniu.

Pasul următor îl constituie analizarea itemilor testului. Sunt eliminaţi acei itemi care nu se referă la comportamente cuprinse în domeniul de conţinut. Pentru itemii rămaşi se identifică zonele domeniului de conţinut pe care le măsoară. La această operaţie este bine să participe cât mai mulţi experţi, care să aprecieze pentru fiecare item dacă se referă la domeniul de conţinut al testului şi dacă este relevant. O analiză atentă o pretind itemii complecşi: aceştia pot să disperseze atenţia subiecţilor şi, prin urmare, scorul lor să fie contaminat de alte caracteristici decât aceea care ar trebui să fie măsurată.

Exemplu.

Li se poate cere elevilor să efectueze aceleaşi operaţii aritmetice formulând două probleme diferite:

a. Care este cel mai mic număr întreg x pentru care are loc inegalitatea x-0.25(x-1) ≥ 5.5 ?

b. Un melc urcă pe un stâlp înalt de 5.5 m, pornind de la bază. În fiecare zi el înaintează cu 1 m, iar în timpul nopţii alunecă în jos cu 0.25 m. În a câta zi melcul atinge vârful stâlpului?

Cea de a doua problemă le propune subiecţilor o sarcină mai complexă decât prima, întrucât le pretinde să citească un text mai lung şi să dea dovadă de atenţie, să observe că întrebarea se referă la ziua în care este "atins" vârful stâlpului. Răspunsul corect este "ziua a şaptea", chiar dacă după ce ajunge în vârful stâlpului, în ziua a şaptea, melcul coboară cu 0.25m în cursul nopţii. Dacă problema a doua este inclusă într-un test de raţionament aritmetic pentru elevi, este posibil ca scorul său să fie alterat de variabila irelevantă "deprinderea de a citi".

Din acest motiv este important să se verifice pentru fiecare item dacă nu cumva necesită un comportament care nu aparţine domeniului de conţinut al testului.

Ultima etapă a validării constă în compararea structurii testului cu structura domeniului de conţinut. Se verifică dacă sunt acoperite toate zonele

domeniului de conţinut şi dacă itemii sunt repartizaţi proporţional cu mărimea şi importanţa zonelor.

Aprecierea generală privind validitatea testului relativă la conţinutul său este subiectivă. Ea nu apelează la metode statistice, ci se bazează doar pe raţionamente. Din acest motiv, validarea relativă la conţinut este denumită în literatură şi *validare logică* sau *validare raţională* (Thorndike, Hagen, 1961; Anastasi, 1976).

Totuşi, în cazul testelor care verifică deprinderi sau cunoştinţe însuşite în cadrul unui program de instruire există recomandarea de a lua în considerare, la aprecierea validităţii relative la conţinut a testului, şi rezultatul comparării statistice a scorurilor obţinute la test de aceleaşi persoane, înainte şi după acest program (Anastasi, 1954). Într-adevăr, dacă există o suprapunere între conţinutul itemilor testului şi conţinutul programului de instruire şi dacă itemii sunt bine construiţi, atunci este de aşteptat ca scorurile finale să fie mult mai mari decât cele iniţiale, deci media diferenţelor de scor să fie pozitivă şi să difere semnificativ de zero.

Validarea relativă la conţinut se impune în cazul testelor utilizate la măsurarea unui atribut ce nu poate fi exprimat printr-un construct (Murphy, Davidshofer, 1991). De exemplu, putem descrie destul de detaliat domeniul de conţinut pentru "cunoştinţe de psihometrie", dar este foarte dificil să găsim comportamentele prin care această caracteristică să poată fi observată.

Validitatea relativă la conţinut se studiază mai ales în cazul testelor de cunoştinţe, atunci când nu există nici un criteriu extern potrivit pentru a analiza validitatea relativă la criteriu.

Trebuie menţionat că validarea relativă la conţinut nu permite să se tragă nici o concluzie despre calitatea măsurătorii realizate de test (Murphy, Davidshofer, 1991). De asemenea, ea nu furnizează suficiente informaţii pentru a putea interpreta corect scorurile testului. Nu se poate afirma că, dacă testul are validitate relativă la conţinutul său, atunci în privinţa atributului măsurat de test, un subiect care a realizat un scor mai mare îi este superior unuia care a obţinut un scor mai mic.

Deci, operaţiile cuprinse sub denumirea de "validarea testului relativă la conţinutul său" nu reprezintă o validare, în sensul definiţiei date de *Standards for Educational and Psychological Tests* din 1985. Ele realizează doar analiza testului în privinţa clarităţii, a reprezentativităţii şi a relevanţei conţinutului său.

9.4. VALIDITATEA TESTULUI RELATIVĂ LA UN CRITERIU

9.4.1. Metode de cercetare a validităţii relative la criteriu

Validarea relativă la criteriu se referă la gradul în care deducţiile făcute pornind de la scorurile testului concordă cu cele bazate pe valorile unei alte măsurări, numită *criteriu*. Validarea relativă la criteriu conţine proceduri care compară scorurile testului cu evaluări, clasificări sau măsurători (mintale ori comportamentale) ale aceloraşi subiecţi.

Pentru fiecare dintre criteriile utilizate la validare, este necesar ca în manualul testului să se explice motivele alegerii sale şi să i se facă o descriere amănunţită (*Standards for Educational and Psychological Tests*, 1985).

Aprecierea validităţii relative la criteriu se bazează, de obicei, pe valorile unui *coeficient de validitate a testului*, care exprimă numeric rezultatul comparării scorurilor testului cu valorile criteriului.

Coeficientul de validitate al unui test, conform definiţiei din majoritatea lucrărilor de teoria testelor, este egal cu coeficientul de corelaţie liniară între scorurile testului şi rezultatele unei măsurători considerate drept variabilă criteriu. Unii autori (Lord, Novick, 1968) înţeleg prin coeficientul de validitate valoarea absolută a coeficientului de corelaţie liniară; alţi autori (Cureton, 1955) impun calculul coeficientului de corelaţie între scorurile observate la test şi scorurile reale ale variabilei criteriu.

În literatura de specialitate există şi recomandarea de a calcula coeficientul de validitate folosind alţi indici în locul coeficientului de corelaţie liniară. De exemplu, Cronbach (1966) permite calcularea validităţii relative la criteriu cu ajutorul coeficientului de corelaţie a rangurilor.

De fapt, în majoritatea cazurilor, testele şi criteriile realizează doar măsurări pe scală ordinală, nu pe scală de interval. Din acest motiv, calculul coeficientului de corelaţie liniară între scorurile testului şi valorile criteriului nu este corectă. Ţinând cont de faptul că prin compararea scorurilor testului cu valorile criteriului se realizează, de fapt, compararea ierarhiilor produse în rândul subiecţilor de test şi de criteriu, pentru studiul validităţii relative la criteriu se pot utiliza toţi coeficienţii care servesc la compararea a două clasamente prezentaţi în Anexa I.

Se apreciază că un test are validitate bună atunci când coeficientul de validitate este mare.

Uneori se poate îmbunătăţi valoarea coeficientului de validitate prin gruparea scorurilor testului, urmată de transformarea valorilor acestora (Albu, Pitariu, 1994 şi 1995a).

O altă posibilitate de cercetare a validităţii unui test relativă la criteriu utilizează două grupe contrastante de subiecţi, formate pe baza valorilor criteriului: una este compusă din indivizi cu rezultate foarte bune, cealaltă din persoane cu rezultate foarte slabe. Se compară statistic mediile scorurilor obţinute la test de persoanele din cele două grupe şi se apreciază că testul este valid relativ la criteriu dacă mediile diferă semnificativ, la un prag de probabilitate fixat (Anastasi, 1976). Această metodă se aplică mai ales atunci când variabila criteriu este calitativă - deci nu se poate calcula coeficientul de corelaţie liniară - şi atunci când criteriul este multidimensional, având mai multe componente cantitative şi/sau calitative.

Unii autori (Hammond, 1995) recomandă ca validarea unui test relativă la criteriu să utilizeze mai multe variabile criteriu, dintre care unele nu au legătură cu ceea ce testul măsoară sau prezice. Acestea servesc la verificarea *specificităţii* testului. Este de dorit ca testul să nu coreleze cu ele.

În cazul testelor utilizate pentru măsurare, validarea relativă la criteriu este o componentă a validării relative la constructul măsurat. Ea apreciază "gradul în care rezultatele furnizate de test corelează cu rezultatele altui instrument de evaluare despre care se presupune că măsoară acelaşi construct sau unul similar" (S. N. Haynes, citat după Silva, 1993). Evident, calculul coeficientului de corelaţie trebuie să fie precedat de analiza constructelor măsurate de test şi de criteriu.

Chiar şi validarea factorială utilizată pentru cercetarea validităţii relative la construct a unui test este, de fapt, o validare relativă la criteriu (Anastasi, 1954). Criteriul este reprezentat, în acest caz, de factorul care corespunde constructului măsurat de test. Una dintre condiţiile impuse pentru a putea aprecia că testul este valid este ca itemii săi să aibă saturaţii mari în factorul respectiv. Or, saturaţia unui item într-un factor este valoarea coeficientului de corelaţie liniară între scorurile itemului şi scorurile factorului.

În cazul testelor utilizate pentru a se lua decizii, validarea relativă la criteriu urmăreşte dacă deciziile luate folosind scorurile testului coincid cu cele care au la bază valorile unei variabile criteriu, despre care se presupune că permite să se ia decizii corecte. De exemplu, decizia poate consta în acceptarea sau respingerea candidaţilor care concurează pentru un post. Se dovedeşte că testul este valid relativ la acest criteriu dacă se constată că indivizii acceptaţi (pe baza unor considerente care nu includ rezultatele testului) sunt tocmai cei care au scoruri mari la test, iar cei respinşi sunt cei cu scoruri mici.

Validarea testelor utilizate pentru predicţie are în vedere gradul în care scorurile testului -numit în acest caz *predictor*- pot fi utilizate pentru a se deduce performanţa la o variabilă numită *criteriu*, independentă de test (Guion, 1974). Trebuie menţionat că o aceeaşi variabilă poate să joace rol de predictor într-un

caz (de exemplu, performanţa şcolară ca predictor pentru performanţa profesională) sau rol de criteriu în alt caz (de exemplu, aceeaşi performanţă şcolară ca şi criteriu pentru un test de inteligenţă).

Există două strategii utilizate în mod obişnuit pentru aprecierea validităţii relative la criteriu a testelor:

- validarea concurentă;
- validarea predictivă.

Validarea concurentă se aplică îndeosebi pentru testele utilizate la obţinerea unor diplome sau licenţe, care măsoară cunoştinţele sau deprinderile persoanelor, şi pentru testele care servesc la stabilirea unui diagnostic clinic. Se cercetează validitatea concurentă atunci când se doreşte să se cunoască în ce măsură scorurile testului estimează poziţiile ocupate în prezent de subiecţi la un criteriu relevant. De exemplu, pentru un test de depresie, validitatea concurentă oferă informaţii despre gradul de coincidenţă dintre concluziile desprinse din interpretarea scorurilor testului şi părerile psihologilor clinicieni sau cele ale psihiatrilor, relative la aceiaşi indivizi.

Pentru cercetarea validităţii concurente este necesară obţinerea scorurilor la criteriu aproximativ în acelaşi timp cu cele ale testului.

Validarea predictivă se aplică atunci când scorurile testului sunt folosite pentru a prezice valorile pe care o anumită variabilă le va avea într-un viitor mai mult sau mai puţin apropiat. Asemenea teste se utilizează de obicei la admiterea în şcoli sau la selecţia profesională, pentru a-i alege pe acei candidaţi care au şanse mari ca la sfârşitul cursurilor sau după ocuparea postului să obţină performanţe bune.

De exemplu, pentru estimarea validităţii predictive a unui test utilizat în selecţia profesională se calculează coeficientul de corelaţie liniară între scorurile testului, administrat cu ocazia finalizării unui curs de pregătire profesională, şi măsurile performanţelor obţinute de aceleaşi persoane după un anumit timp, în care au avut posibilitatea să aplice cunoştinţele respective.

Trebuie subliniat faptul că validitatea predictivă se cercetează folosind loturi neselecţionate. Aceasta înseamnă că toate persoanele cărora li s-a administrat testul vor urma acelaşi tratament şi vor fi prezente la evaluarea variabilei criteriu. În acest fel se asigură ca în lotul de subiecţi care serveşte la compararea testului cu criteriul să existe atât scoruri mari la test, cât şi scoruri mici. Nu totdeauna se poate realiza aşa ceva. Este greu de acceptat, de exemplu, ca să fie angajate toate persoanele care se prezintă la selecţia profesională sau ca decizia referitoare la angajare să se ia la întâmplare.

Atunci când testul va fi utilizat pentru a face predicţii, este necesar ca, alături de coeficientul de validitate al testului să se determine şi ecuaţia de regresie

$$\text{valoarea criteriului} = f(\text{scor la test})$$

Aceasta permite să se estimeze rezultatul pe care îl va obţine la criteriu fiecare persoană, cunoscând scorul său la test. De exemplu, dacă relaţia dintre media de absolvire a facultăţii, F, şi media obţinută la admiterea în facultate, A, poate fi scrisă sub forma

$$F = 0.3 + 0.9A$$

atunci este de aşteptat ca o persoană care a fost admisă în facultate cu media 8.5 să termine facultatea cu media

$$F = 0.3 + 0.9 \times 8.5 = 7.95$$

9.4.2. Problema restrângerii mulţimii

Deosebirea esenţială dintre cele două tipuri de validări relative la criteriu nu constă în lungimea intervalului de timp dintre administrarea testului şi înregistrarea rezultatelor criteriului, ci în faptul că validarea concurentă utilizează un eşantion de populaţie selecţionat (personal angajat deja sau studenţi admişi în facultăţi). Apare aici următoarea problemă, denumită *restrângerea mulţimii*, care face ca validarea concurentă să nu poată furniza totdeauna informaţii corecte asupra capacităţii predictive a testului:

Lotul de persoane utilizat la calculul coeficientului de validitate provine din populaţia asupra căreia ar fi trebuit să se facă predicţia, dar până în momentul când se înregistrează valorile variabilei criteriu el a fost supus deja la două selecţii.

(1) Pe de o parte, au fost reţinuţi din eşantionul iniţial doar cei care au obţinut rezultate bune la criteriu (de exemplu, cei care au avut note mari la admiterea în facultate). Dacă acest criteriu măsoară acelaşi lucru ca şi testul, atunci persoanele rămase sunt cele care vor obţine la test scoruri mai mari decât ar fi realizat cei respinşi.

(2) Pe de altă parte, au fost eliminaţi pe parcurs cei cu performanţe şcolare sau profesionale slabe, adică cei cu valori mici la variabila criteriu utilizată în studiul validităţii. Este posibil să fi fost înlăturaţi prin promovare, şi cei cu valori foarte mari la criteriu.

Deci, pentru lotul considerat, atât mulţimea scorurilor testului cât şi mulţimea valorilor criteriului reprezintă restrângeri ale mulţimilor corespunzătoare care s-ar fi obţinut pentru întreaga populaţie. În consecinţă, coeficientul de corelaţie liniară calculat pe baza persoanelor selecţionate diferă de cel care ar fi rezultat lucrând cu întreaga populaţie şi care ar fi fost util pentru aprecierea valorii predictive a testului.

Exemplu.

Să considerăm un lot format din 105 persoane care au răspuns la un test de capacitate de organizare, ale cărui scoruri posibile sunt 1, 2, 3, 4 şi 5. Fiecare subiect a fost evaluat în privinţa capacităţilor sale manageriale, primind o notă cuprinsă între 1 şi 5.

Rezultatele obţinute de persoanele din lot sunt notate în următorul tabel de frecvenţe:

Scorul testului	Valoarea criteriului				
	1	2	3	4	5
1	10	2	2	1	2
2	2	8	4	0	2
3	0	6	18	3	0
4	2	1	4	11	6
5	1	2	4	7	7

Pe întregul lot, coeficientul de corelaţie liniară între test şi criteriu are valoarea 0.520, semnificativă la pragul p=0.001.

Dacă din lot au fost reţinuţi doar cei 45 de subiecţi care au realizat la test cel puţin 4 puncte, atunci coeficientul de corelaţie liniară între test şi criteriu este egal cu 0.026, fiind nesemnificativ la pragul p=0.05.

Dacă din lot au fost eliminate persoanele care au fost evaluate cu note mai mici decât 4, atunci pentru cele 39 de persoane rămase coeficientul de corelaţie liniară dintre test şi criteriu ia valoarea -0.095, fiind nesemnificativ la pragul p=0.05.

Pentru 31 de persoane care au primit cel puţin nota 4 la evaluare şi au obţinut cel puţin 4 puncte la test coeficientul de corelaţie între test şi criteriu este egal cu 0.148, fiind nesemnificativ la pragul p=0.05.

9.4.3. Interpretarea coeficienţilor de validitate

Hull (Guilford, 1965) a impus condiţia ca un test să fie utilizat în practică doar dacă are un coeficient de validitate mai mare decât 0.45. Această cerinţă este însă greu de realizat. Un test bun rareori poate da, cu un criteriu

important, un coeficient de corelaţie mai mare decât 0.5 (Murphy, Davidshofer, 1991).

În experienţa obişnuită, valoarea coeficientului de validitate al testelor utilizate pentru a face predicţii este cuprinsă între 0 şi 0.60, în majoritatea cazurilor fiind situată în partea inferioară a intervalului [0, 0.60] (Guilford, 1965).

Constatarea că, în general, comportamentul aceleiaşi persoane variază de la o situaţie la alta a condus la concluzia că valoarea coeficientului de validitate al testelor de personalitate administrate în scopul predicţiei unei acţiuni nu poate fi mai mare decât 0.4 (Funder, 1983).

O posibilitate de a stabili dacă validitatea relativă la criteriu a unui test este sau nu acceptabilă porneşte de la eroarea maximă acceptată în estimarea valorilor criteriului şi se bazează pe formula de calcul a erorii standard de estimare (Gregory, 1992):

$$ES_{est}=\sigma_c\sqrt{1-r_{ct}^2} \quad (9.4.3.1)$$

unde:

ES_{est} = eroarea standard de estimare, care reprezintă marginea superioară a erorii de estimare a valorilor criteriului prin ecuaţia de regresie liniară;
σ_c = abaterea standard a valorilor criteriului;
r_{ct} = coeficientul de corelaţie liniară între test şi criteriu;

ES_{est} ajută să se răspundă la întrebarea "Cât de precis poate fi prezisă performanţa la criteriu, pornind de la scorurile testului?"

Presupunând că repartiţia de frecvenţe a valorilor criteriului urmează legea normală şi că ecuaţia de regresie a criteriului (c) faţă de test (t) este:

$c = a + bt$

există şanse de 95% ca valoarea criteriului pentru o persoană care a obţinut scorul t_0 la test să se afle în intervalul:

$[a+bt_0-1.96ES_{est}, a+bt_0+1.96ES_{est}]$

Cu cât coeficientul de validitate al testului, r_{ct}, este mai mare, cu atât eroarea standard de estimare este mai mică, iar predicţia realizată cu ajutorul testului este mai precisă.

9.4.4. Factori care afectează valoarea coeficienţilor de validitate

Valoarea coeficientului de validitate al unui test este influenţată nu numai de test, ci şi de criteriul utilizat şi de populaţia căreia i se administrează

testul. Amintim câţiva dintre factorii care afectează validitatea testului (adaptare după Aiken, 1994 şi Warr, 1997).

A. Unele caracteristici ale populaţiei, precum sexul, vârsta sau chiar anumite trăsături de personalitate, acţionează ca variabile moderatoare, modificând relaţia dintre criteriu şi test. Subiecţii se împart în grupe, în funcţie de valorile variabilei moderatoare, iar valoarea coeficientului de validitate al testului se poate schimba de la o grupă la alta.

În general, coeficienţii de validitate sunt mai mici în grupele mai omogene, adică în grupele în care scorurile testului sau valorile criteriului variază puţin.

Atunci când valoarea coeficientului de validitate al unui test este scăzută, iar ceea ce se cunoaşte despre test şi despre criteriu nu oferă o explicaţie, trebuie să se verifice dacă nu există o variabilă moderatoare care se interpune între criteriu şi test.

Întrucât validitatea unui test este dependentă şi de grupul de subiecţi, ori de câte ori se constată că un test nou este valid faţă de un anumit criteriu, este necesar să se efectueze o validare suplimentară (cross-validation), folosind acelaşi criteriu, dar un alt eşantion de subiecţi, proveniţi din aceeaşi populaţie ca şi primul. Doar dacă şi în acest caz testul se dovedeşte a fi valid, se poate recomanda utilizarea sa în practică.

B. La fel ca şi fidelitatea, validitatea relativă la criteriu este influenţată de lungimea testului.

Dacă testul este format din itemi paraleli şi se modifică lungimea testului prin eliminarea unor itemi sau prin adăugarea de itemi paraleli, atunci coeficientul de validitate al testului obţinut, relativ la acelaşi criteriu pentru care a fost calculat coeficientul de validitate al testului original, este dat de formula (Burisch, 1997):

$$r_k = \frac{r_0\sqrt{k}}{\sqrt{1+(k-1)r_{tt}}} \qquad (9.4.4.1)$$

unde:

r_k = coeficientul de validitate al testului nou;
r_0 = coeficientul de validitate al testului original;
k = raportul dintre numărul de itemi în testul nou şi numărul de itemi în testul original;
r_{tt} = coeficientul de fidelitate al testului original.

Se constată că prin adăugarea de itemi paraleli la un test validitatea testului creşte.

C. Validitatea testului este dependentă şi de validitatea criteriului ca şi măsură a variabilei la care se referă testul.

Dacă valorile variabilei criteriu sunt influenţate de scorurile testului, se spune că a avut loc o *contaminare a criteriului*. Aceasta are ca efect creşterea artificială a validităţii relative la criteriu a testului.

În majoritatea cazurilor, contaminarea criteriului se produce atunci când testul şi criteriul au itemi comuni sau în acele situaţii când valorile criteriului sunt rezultatul unor evaluări făcute de experţi, iar aceştia cunosc scorurile obţinute de către subiecţi la test şi, intenţionat sau nu, sunt influenţaţi de ele în formularea aprecierilor.

D. Un alt factor care influenţează mărimea coeficientului de validitate este frecvenţa relativă în populaţie a persoanelor care prezintă caracteristica la care se referă testul (base rate). Atunci când comportamentul pe care testul îl măsoară sau îl prezice este foarte rar întâlnit (precum psihoticismul sau suicidul), validitatea relativă la criteriu este de obicei scăzută.

E. Validitatea unui test este dependentă de fidelitatea sa. Într-adevăr, dacă $a \in M$ şi $b \in M$ sunt două măsurători paralele, iar $c \in M$ este o măsurătoare arbitrară, atunci are loc următoarea inegalitate:

$$\rho(X_{a*}, X_{c*}) \leq \sqrt{\rho(X_{a*}, X_{b*})} \qquad (9.4.4.2)$$

Aceasta se interpretează astfel: coeficientul de validitate al unui test $a \in M$ relativ la orice criteriu $c \in M$ nu poate depăşi indicele de fidelitate al testului a. Înseamnă că, dacă un test nu este fidel, validitatea sa relativ la orice criteriu nu poate fi mare. Dar, o valoare mare a indicelui de fidelitate nu garantează o validitate mare.

F. Uneori, coeficientul de validitate ia o valoare scăzută nu din cauza testului, ci a criteriului utilizat. Aşa se întâmplă atunci când:
- criteriul are o fidelitate scăzută;
- intervalului de valori pentru criteriu este restrâns, neacoperind decât o parte din mulţimea valorilor posibile;
- repartiţia de frecvenţe a valorilor criteriului nu este normală;

• valorile criteriului nu sunt acordate sau notate corect, fie din cauza unor factori care ţin de evaluator (oboseală, neatenţie, incompetenţă sau părtinire), fie pentru că respectivul criteriu este dificil de măsurat.

9.5. CÂTEVA IDEI IMPORTANTE DESPRE VALIDITATEA TESTELOR

La alegerea strategiilor de validare a unui test, precum şi la interpretarea coeficienţilor de validitate, este util să se ia în considerare şi următoarele păreri (adaptare după Messick, 1981 şi Hammond, 1995):

A. Validitatea unui test este dependentă de scopul pentru care acesta este folosit şi de populaţia căreia i se administrează testul. Nici un test nu este valid în orice situaţie de utilizare. Concluziile desprinse dintr-un studiu privind validitatea unui test sunt influenţate şi de metodele aplicate în cercetare.

B. Există diverse metode pentru validarea testelor: analiza de itemi, analiza factorială, calculul coeficienţilor de corelaţie dintre test şi diverse criterii etc. Alegerea strategiei de validare pentru un anumit test trebuie să fie determinată de scopul pentru care se va utiliza testul şi de cerinţele formulate de cel care a solicitat examinarea psihologică în cadrul căreia se intenţionează să se administreze testul respectiv.

C. Validitatea unui test nu poate fi estimată printr-un singur coeficient, ci se deduce din acumularea dovezilor empirice şi conceptuale.

Standards for Educational and Psychological Tests din 1974 atrag atenţia că "validitatea nu se măsoară, ci se deduce". Manualul testului trebuie să conţină valorile mai multor coeficienţi de validitate, calculaţi pentru utilizări diferite ale testului sau pentru populaţii diferite. În cazul unei aplicări particulare a testului nu se va lua în considerare un singur coeficient, ci prin analiza lor globală se va aprecia dacă utilizarea testului este "adecvată" sau "nesatisfăcătoare".

Pentru a sublinia ideea că este greşit să se spună "testul are validitate (de un anumit tip)", *Standards for Educational and Psychological Tests* din 1985 modifică terminologia utilizată în legătură cu validitatea. Astfel, ele înlocuiesc denumirile cunoscute ale "tipurilor" de validitate cu următoarele "etichete" atribuite strategiilor de validare: "*dovezi ale validităţii relative la conţinut*", "*dovezi ale validităţii relative la construct*" şi "*dovezi ale validităţii*

relative la criteriu". Prin urmare, faptul că scorurile testului folosit ca predictor corelează liniar semnificativ, la un prag de probabilitate fixat, cu valorile unui criteriu nu reprezintă o evaluare a validităţii relative la criteriu şi nu permite să se afirme că "testul are validitate relativă la criteriu", ci doar constituie una dintre dovezile validităţii relative la criteriu a testului.

D. Strategiile de validare se aplică încă din faza de construcţie a testului. Ele se repetă până când se obţine un test acceptabil în ceea ce priveşte lungimea, fidelitatea, validitatea şi uşurinţa în administrare, cotare şi interpretare. Procedeele de validare sunt următoarele:

a. În cazul testelor care măsoară constructe, se aleg şi se introduc în test itemi din două categorii: unii consideraţi a fi reprezentativi pentru domeniul constructului şi unii despre care se presupune că nu au nici o legătură cu constructul. După administrarea testului sunt reţinuţi doar acei itemi care se aseamănă între ei în privinţa răspunsurilor obţinute de la subiecţi. Se analizează conţinutul acestora pentru a se vedea dacă, într-adevăr, ei exemplifică teoria constructului ce trebuie măsurat şi dacă se deosebesc de itemii consideraţi a fi nerelevanţi pentru construct.

b. În cazul testelor care vor fi validate relativ la conţinutul lor, se introduc în test doar itemi care acoperă domeniul de conţinut al testului. După administrarea testului, se reţin itemii la care s-au obţinut răspunsuri asemănătoare. În măsura posibilităţii, se verifică dacă ei corelează cu un criteriu care se referă la acelaşi domeniu de conţinut sau dacă răspunsurile obţinute la aceşti itemi se deosebesc semnificativ între două grupe contrastante de subiecţi.

c. În cazul testelor care vor fi utilizate pentru a face predicţii sau a lua decizii, pentru care se poate alege un criteriu de validare, după administrarea testului se reţin doar itemii care se comportă faţă de acest criteriu aşa cum se doreşte să se comporte întregul test.

E. Validarea unui test este un proces care trebuie repet mereu pe parcursul perioadei în care acesta este utilizat, întrucât datele care se acumulează, provenite din observaţii clinice şi din diverse proiecte de cercetare, permit ca interpretarea propusă iniţial pentru scorurile testului să fie îmbunătăţită şi rafinată.

Ori de câte ori se face o modificare importantă într-un test în ceea ce priveşte formatul, limbajul, conţinutul sau instrucţiunile de administrare, testul va fi validat din nou.

CAPITOLUL 10
DIFICULTATEA ŞI SENSIBILITATEA UNUI TEST PSIHOLOGIC

Aceste proprietăţi sunt discutate, de obicei, atunci când se verifică dacă un test este adecvat unui lot de persoane. Elementele analizate în acest scop sunt, în majoritatea cazurilor, forma repartiţiei de frecvenţe şi valorile indicilor statistici (media, dispersia) pentru scorurile realizate la test de lotul respectiv sau de un eşantion asemănător acestuia.

10.1. DIFICULTATEA UNUI TEST PSIHOLOGIC

Se consideră că un test este **dificil** dacă un număr mare de persoane obţin scoruri mici, este **uşor** dacă majoritatea scorurilor realizate sunt mari şi este **potrivit** lotului dacă scorurile sunt repartizate pe un interval de valori destul de mare, de obicei fiind grupate în jurul mediei. În primul caz, repartiţia de frecvenţe pentru scorurile obţinute la test este asimetrică la stânga (are forma literei *i*) (figura 10.1.1a), în cel de-al doilea caz este asimetrică la dreapta (are forma literei *j*) (figura 10.1.1.b), iar în ultimul caz este simetrică faţă de verticala ridicată în punctul corespunzător mediei scorurilor (figura 10.1.1c). De obicei se apreciază că un test este *adecvat unei populaţii* dacă repartiţia scorurilor sale urmează legea normală.

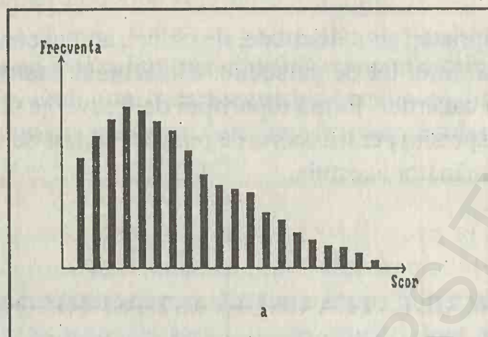
O asemenea analiză a dificultăţii este posibilă nu numai pentru testele de eficienţă (de aptitudini şi de cunoştinţe), ci şi pentru chestionarele de personalitate, deşi în cazul acestora din urmă termenul "dificultate" nu pare potrivit.

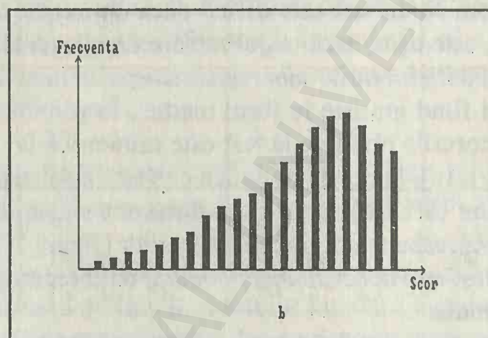
Dificultatea unui test, apreciată pe baza repartiţiei de frecvenţe a scorurilor rezultate în urma unei administrări, este determinată atât de conţinutul testului, cât şi de caracteristicile lotului de subiecţi.

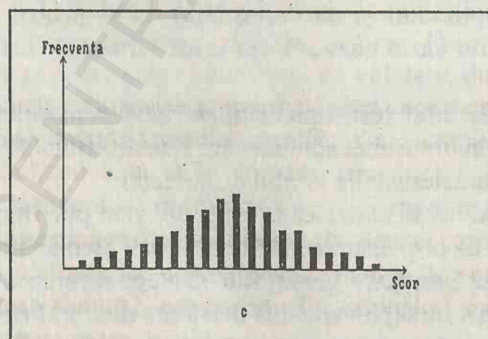
Atunci când itemii care alcătuiesc testul sunt puternic corelaţi între ei, fiecare persoană va da răspunsuri cotate identic sau asemănător, pentru că îi va rezolva pe toţi corect (respectiv, greşit) sau va alege răspunsul care corespunde valorii mari (respectiv, mici) a variabilei măsurate de test. Prin urmare, în orice lot de indivizi se vor obţine doar scoruri foarte mari şi/sau foarte mici.

Pentru un chestionar de personalitate cu itemi dihotomici este posibil ca populaţia căreia i se administrează testul să se împartă în două grupe: unii care la majoritatea itemilor dau răspunsul corespunzător scorului 1 şi alţii care

aleg răspunsul cotat cu 0. Se obţine, deci, o repartiţie bimodală a scorurilor (figura 10.1.1.d). O asemenea situaţie poate să apară, de pildă, la o scală de feminitate/ masculinitate, când răspunsurile la itemi se diferenţiază în funcţie de sex.







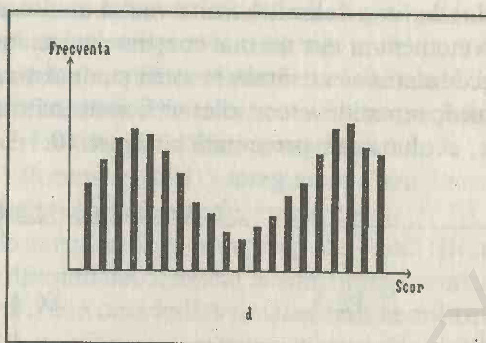


Figura 10.1.1.
Diverse repartiţii de frecvenţe ale scorurilor unor teste

Testele centrate pe criteriu produc, de obicei, repartiţii asimetrice sau bimodale. Ele conţin itemi care verifică existenţa unor priceperi sau cunoştinţe absolut necesare unui anumit scop (promovarea unui program de instruire, ocuparea unui post etc.). De exemplu, ele pot solicita rezolvarea unor exerciţii de aritmetică simple, la sfârşitul clasei întâi, sau executarea unei piese dificile, pentru ocuparea unui post care pretinde o calificare superioară. Utilizarea acestor teste nu urmăreşte ierarhizarea persoanelor, ci împărţirea lor în două categorii: "capabile" şi "incapabile" pentru scopul fixat.

Un test de cunoştinţe, administrat aceloraşi indivizi în momente diferite, poate produce repartiţii total deosebite ale scorurilor.

Exemplu (adaptare după G. de Landsheere, 1975).

Să considerăm un test de cunoştinţe de statistică, destinat studenţilor anului I de la Psihologie.

La începutul anului, majoritatea studenţilor nu cunosc principalele formule statistice utilizate la prelucrarea datelor experimentale; unii ştiu să calculeze media, mediana şi dispersia şi doar câţiva au lucrat cu coeficienţi de corelaţie. Rezultatele care s-ar obţine la test ar avea curba de repartiţie asimetrică la stânga.

După trecerea examenului de "Psihologie experimentală şi analiza datelor", majoritatea studenţilor cunosc şi aplică bine formulele statistice; unii au dificultăţi la aplicarea câtorva formule şi numai foarte puţini nu cunosc formulele. Acelaşi test de cunoştinţe ar avea curba de repartiţie a scorurilor asimetrică la dreapta.

În intervalul de timp delimitat de începutul anului şi sfârşitul sesiunii există, probabil, un moment în care cei mai conştiincioşi au învăţat mai mult, cei mai puţin interesaţi de statistică au rămas în urmă şi, dacă s-ar administra testul de cunoştinţe, curba de repartiţie a scorurilor ar fi asemănătoare curbei normale.

Schematic, evoluţia este prezentată în figura 10.1.2.

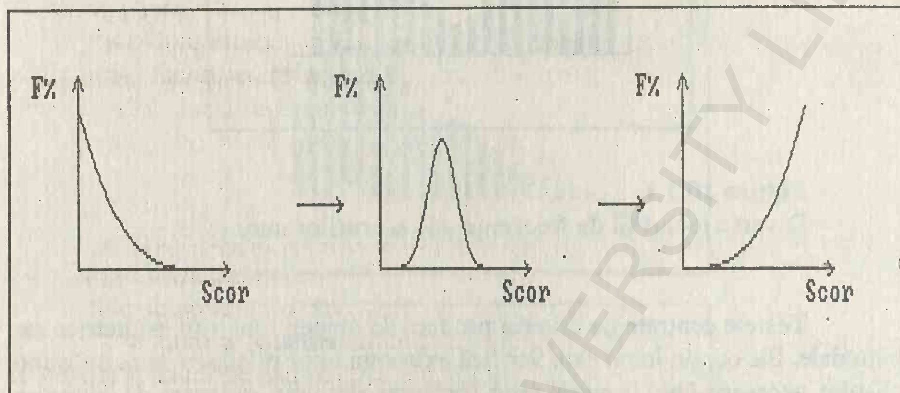


Figura 10.1.2
Curbele de repartiţie pentru scorurile aceluiaşi test de cunoştinţe, în trei momente diferite

10.2. SENSIBILITATEA UNUI TEST PSIHOLOGIC

Puterea de discriminare sau **sensibilitatea** unui test este dată de capacitatea acestuia de a face distincţie între persoanele examinate, în privinţa trăsăturii de personalitate sau a aptitudinii măsurate (Lafon, 1973), respectiv de a produce scoruri diferite pentru subiecţii care se deosebesc între ei în privinţa caracteristicii măsurate de test (Kline, 1993).

Piéron (1952) consideră că puterea de discriminare a unui test este dată de numărul claselor pe care le permite în cadrul unui grup. Cu cât numărul valorilor de scor obţinute de persoanele unui lot este mai mare, cu atât testul diferenţiază mai bine între indivizi. Dar, numărul de valori de scor rezultate în urma administrării testului depinde atât de test (de itemii componenţi), cât şi de caracteristicile lotului (cât de mari sunt deosebirile dintre persoane în privinţa variabilei măsurate sau prezise de test).

Atunci când cota brută a testului se obţine prin însumarea cotelor itemilor, este de aşteptat ca un test alcătuit din mulţi itemi să furnizeze mai multe valori de scor decât unul format dintr-un număr redus de itemi. Dar, aşa cum s-a arătat mai sus, dacă itemii sunt puternic intercorelaţi, există riscul de a se obţine doar scoruri foarte mari şi/sau foarte mici.

Ausubel şi Robinson (1981) atrag atenţia asupra dependenţei dintre sensibilitatea testului şi forma repartiţiei scorurilor.

O repartiţie a rezultatelor apropiată ca formă de curba normală, de exemplu, asigură o discriminare maximă la ambele capete ale scalei (unde sunt risipite puţine scoruri) şi o discriminare mai redusă la mijlocul scalei (unde se îngrămădesc laolaltă numeroase rezultate); o repartiţie uniformă a scorurilor asigură o discriminare la fel de bună de-a lungul întregii scale. O repartiţie asimetrică (în care un număr disproporţionat de mare de rezultate se acumulează la un capăt al scalei) este extrem de discriminativă la capătul unde găsim puţine scoruri.

Se deduce, deci, că un test discriminează cu atât mai bine în cadrul întregului lot de subiecţi, cu cât permite un număr mai mare de valori de scor şi cu cât repartiţia acestora este mai uniformă.

Pentru un test alcătuit din itemi dihotomici, cotaţi cu 0 şi 1, al cărui scor se obţine prin însumarea cotelor itemilor, o măsură a puterii de discriminare bazată pe forma repartiţiei scorurilor obţinute la test, o reprezintă coeficientul lui Ferguson (Kline, 1993). Acesta se calculează cu ajutorul formulei:

$$\delta = \frac{n+1}{nN^2}\left(N^2 - \sum_{i=0}^{n} f_i^2\right) \qquad (10.2.1)$$

unde:

δ = coeficientul lui Ferguson;
n = numărul itemilor testului;
N = numărul de subiecţi cărora li s-a administrat testul;
f_i = frecvenţa scorului i, i=0,1,...,n.

Coeficientul lui Ferguson ia valori între 0 (atunci când apare un singur scor) şi 1 (în cazul repartiţiei uniforme). Repartiţia legii normale furnizează un coeficient egal cu 0.93. Se consideră că un test are o capacitate de discriminare bună dacă δ ia o valoare mai mare decât 0.9.

Există o legătură puternică între dificultatea unui test şi sensibilitatea sa. Atunci când testul este foarte greu sau foarte uşor pentru un grup de persoane, rezultatele se concentrează spre una dintre extremităţile intervalului

de valori posibile pentru scoruri şi, deci, nu sunt evidenţiate deosebirile dintre indivizi.

Se impune o observaţie: forma repartiţiei de frecvenţe a scorurilor este dependentă atât de forma repartiţiei de frecvenţe a valorilor variabilei măsurate, în grupul testat, cât şi de test (de caracteristicile psihometrice ale itemilor acestuia).

Exemplu (Albu, Pitariu, 1992a).

S-a considerat un test alcătuit din patru itemi, ale căror funcţii caracteristice au expresia analitică:

$$P(y) = c + \frac{1-c}{1+e^{-1,7a(y-b)}}$$

S-a presupus că pentru o grupă de subiecţi variabila măsurată de test are o repartiţie simetrică.

Cu ajutorul calculatorului s-a determinat repartiţia de frecvenţe a scorurilor testului pentru diverse alegeri ale parametrilor *a*, *b* şi *c*. În figura 10.2.1 sunt prezentate două dintre repartiţiile de frecvenţe obţinute. Se constată că între forma repartiţiei variabilei măsurate în eşantionul de subiecţi şi forma repartiţiei scorurilor realizate pot exista deosebiri importante.

Atunci când repartiţia de frecvenţe a scorurilor obţinute are o asimetrie puternică, dificultatea şi sensibilitatea testului pot fi analizate identificând scorurile care apar mai frecvent. În celelalte cazuri, însă, pentru a putea aprecia corect aceste proprietăţi ale testului este necesar să se calculeze principalii indicatori statistici ai scorurilor şi să se compare, folosind teste statistice, valorile lor cu cele înregistrate în manualul testului sau obţinute în alte ocazii.

Media rezultatelor aduce informaţii despre nivelul general al variabilei măsurate de test. Dacă media grupului examinat este semnificativ mai scăzută (respectiv, mai ridicată) decât cea a populaţiei care a servit la etalonarea testului se deduce că testul este prea dificil (respectiv, prea uşor) pentru lotul respectiv şi interpretarea scorului unei persoane prin raportare la etalonul din manual ar putea conduce la concluzii greşite.

Abaterea standard serveşte la aprecierea omogenităţii grupului examinat. O valoare mică a abaterii standard este caracteristică unui lot de subiecţi foarte asemănători între ei în privinţa variabilei măsurate de test. În acest caz, testul nu este sensibil, nu realizează o diferenţiere bună între persoane. Valorile mari ale abaterii standard se întâlnesc în loturile neomogene, cu deosebiri mari între cei testaţi. Dar atunci media scorurilor este mai puţin

ilustrativă pentru nivelul lotului, întrucât rezultatele obţinute sunt împrăştiate într-un interval destul de mare în jurul valorii medii.

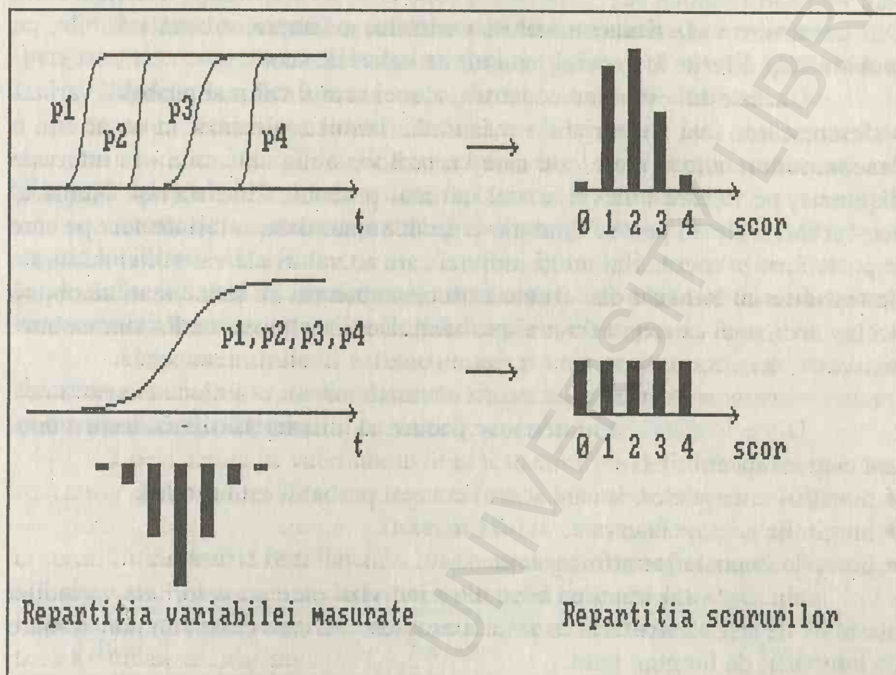


Figura 10.2.1
Comparaţie între repartiţia valorilor variabilei măsurate de test şi repartiţia scorurilor testului

Trebuie observat că nici un test nu poate fi atât de sensibil încât totdeauna două persoane deosebite în privinţa variabilei măsurate de test să obţină scoruri diferite. Aceasta, în primul rând, deoarece trăsăturile psihice, aptitudinile, cunoştinţele şi celelalte caracteristici măsurate de teste sunt variabile continue, în timp ce scorurile testelor sunt variabile discrete, având doar un număr finit de valori. Prin urmare, nici un test nu poate stabili o corespondenţă biunivocă între variabila măsurată şi scorurile pe care le produce. Oricât de deosebiţi între ei ar fi indivizii dintr-un lot examinat, dacă numărul lor este mai mare decât numărul de valori de scor pe care le poate furniza testul, vor exista persoane cu acelaşi scor, deci între care testul nu diferenţiază.

Pe de altă parte, scorul realizat de o persoană la un test este afectat de diverse erori (erori cauzate de starea de oboseală a subiectului, de condiţiile în care s-a făcut examinarea şi cotarea sau chiar erori de construcţie a testului etc.). Din acest motiv, la fiecare nivel al variabilei măsurate, t, sunt posibile, cu probabilităţi diferite în general, mai multe valori de scor.

Dacă testul este bine construit, atunci scorul cel mai probabil variază nedescrescător faţă de variabila măsurată. Testul realizează în acest caz o descompunere a axei reale - pe care variază variabila măsurată - în intervale disjuncte, pe fiecare interval scorul cel mai probabil fiind acelaşi. Numărul acestor intervale nu poate fi mai mare decât numărul de valori de scor pe care le poate furniza testul. Mai mulţi indivizi care au valori ale variabilei măsurate de test diferite între ele dar situate în acelaşi interval au şanse mari să obţină acelaşi scor, egal cu scorul cel mai probabil. Deci, testul nu va discrimina între ei.

Din cele arătate până acum se deduce că puterea de discriminare a unui test este legată de:

- numărul intervalelor pe care scorul cel mai probabil este acelaşi;
- lungimile acestor intervale;
- poziţiile intervalelor pe axa reală.

Un test va discrimina bine între indivizii care au valori ale variabilei măsurate de test plasate în acea zonă a axei reale în care există un număr mare de intervale, de lungimi mici.

10.3. INFLUENŢA CARACTERISTICILOR ITEMILOR ASUPRA DIFICULTĂŢII ŞI A SENSIBILITĂŢII TESTULUI

În continuare vom arăta, prin câteva exemple, cum depind dificultatea şi sensibilitatea unui test de caracteristicile itemilor componenţi.

Ne referim doar la teste alcătuite din itemi care măsoară aceeaşi variabilă, dihotomici, cotaţi cu 1 (pentru valori mari ale variabilei) şi 0 (în cazul valorilor mici ale variabilei).

Pentru un asemenea item, indicele de dificultate este dat de frecvenţa relativă a persoanelor care au obţinut scorul 1 în mulţimea subiecţilor care au răspuns la item. El arată cât de dificil este itemul pentru lotul de persoane testat.

În toate exemplele analizate itemii sunt valizi, în sensul că pentru fiecare item scorul 0 este obţinut de persoane care la variabila măsurată de test au valori mai mici decât persoanele care obţin la item scorul 1.

Să considerăm 10 persoane $s_1, s_2,...,s_{10}$ care au răspuns la 6 itemi, $i_1, i_2,...,i_6$. Scorul testului format din aceşti itemi se obţine prin însumarea scorurilor itemilor.

Notăm cu $c_1, c_2,..., c_{10}$ valorile variabilei măsurate de test pentru cele 10 persoane şi presupunem că între ele există relaţiile:

$$c_1 < c_2 < ... < c_{10} \quad (10.3.1)$$

Exemplul 10.3.1

Scorurile obţinute de subiecţi la itemi şi la testul compus din ei sunt notate în Tabelul 10.3.1.

Se constată că dacă itemii componenţi au aceeaşi dificultate şi sunt intercorelaţi, testul nu discriminează bine între subiecţi.

Alegerea unor itemi valizi cu aceeaşi valoare a indicelui de dificultate micşorează numărul scorurilor distincte obţinute la test. Acest lucru se observă şi din următorul exemplu.

Tabelul 10.3.1

Scorurile itemilor $i_1, i_2,...,i_6$ pentru subiecţii $s_1, s_2,...,s_{10}$

Item	Persoana										Dificultate
	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	
i_1	0	0	0	0	0	1	1	1	1	1	0.5
i_2	0	0	0	0	0	1	1	1	1	1	0.5
i_3	0	0	0	0	0	1	1	1	1	1	0.5
i_4	0	0	0	0	0	1	1	1	1	1	0.5
i_5	0	0	0	0	0	1	1	1	1	1	0.5
i_6	0	0	0	0	0	1	1	1	1	1	0.5
Scor total	0	0	0	0	0	6	6	6	6	6	

Exemplul 10.3.2

Scorurile obţinute de subiecţi la itemi şi la testul compus din ei sunt notate în Tabelul 10.3.2.

Tabelul 10.3.2
Scorurile itemilor $i_1, i_2,...,i_6$ pentru subiecţii $s_1, s_2,...,s_{10}$

Item	Persoana										Dificultate
	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	
i_1	0	0	0	0	0	0	1	1	1	1	0.4
i_2	0	0	0	0	0	0	1	1	1	1	0.4
i_3	0	0	0	0	0	0	1	1	1	1	0.4
i_4	0	0	0	0	1	1	1	1	1	1	0.6
i_5	0	0	0	0	1	1	1	1	1	1	0.6
i_6	0	0	0	0	1	1	1	1	1	1	0.6
Scor total	0	0	0	0	3	3	6	6	6	6	

Din cele 7 valori posibile de scor la test s-au obţinut doar 3. Scorurile testului nu discriminează între subiecţii cu valori mari ale variabilei măsurate (care obţin, toţi, scorul maxim 6) şi nici între cei cu valori mici ale acestei variabile (care obţin, toţi, scorul minim, 0). Cauza o constituie faptul că toţi itemii componenţi corelează între ei şi sunt de dificultate medie.

Exemplul 10.3.3

Scorurile obţinute de subiecţi la itemi şi la testul compus din ei sunt notate în Tabelul 10.3.3.

Din cele 7 valori posibile de scor la test s-au obţinut 5: 0, 1, 3, 5 şi 6. Scorurile testului sunt repartizate aproximativ echidistant în intervalul scorurilor posibile, [0,6], datorită faptului că şi dificultăţile itemilor care compun testul sunt aproximativ echidistante în intervalul [0,1]. Testul format este de dificultate medie.

Tabelul 10.3.3

Scorurile itemilor $i_1, i_2, \ldots, i_6$ pentru subiecţii $s_1, s_2, \ldots, s_{10}$

Item	Persoana										Dificultate
	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	
i_1	0	0	0	0	0	0	0	0	1	1	0.2
i_2	0	0	0	0	0	0	1	1	1	1	0.4
i_3	0	0	0	0	0	0	1	1	1	1	0.4
i_4	0	0	0	0	1	1	1	1	1	1	0.6
i_5	0	0	0	0	1	1	1	1	1	1	0.6
i_6	0	0	1	1	1	1	1	1	1	1	0.8
Scor total	0	0	1	1	3	3	5	5	6	6	

Cu totul alta este situaţia în următorul exemplu.

Exemplul 10.3.4

Tabelul 10.3.4

Scorurile itemilor $i_1, i_2, \ldots, i_6$ pentru subiecţii $s_1, s_2, \ldots, s_{10}$

Item	Persoana										Dificultate
	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	
i_1	0	0	0	0	0	0	0	1	1	1	0.3
i_2	0	0	0	0	0	1	1	1	1	1	0.5
i_3	0	0	0	0	1	1	1	1	1	1	0.6
i_4	0	0	0	1	1	1	1	1	1	1	0.7
i_5	0	0	0	1	1	1	1	1	1	1	0.7
i_6	0	0	0	1	1	1	1	1	1	1	0.7
Scor total	0	0	0	3	4	5	5	6	6	6	

Scorurile obţinute de subiecţi la itemi şi la testul compus din ei sunt notate în Tabelul 10.3.4. Şi în acest caz s-au obţinut tot 5 scoruri la test, dar acestea sunt mai apropiate de scorul maxim posibil decât în exemplul precedent, testul este mai uşor. Aceasta, din cauza faptului că itemii care compun testul au indicele de dificultate mai mare decât în exemplul precedent.

Pentru a obţine cât mai multe valori de scor se aleg itemi care să aibă indici de dificultate diferiţi şi repartizaţi cât mai uniform pe intervalul [0,1], ca în exemplul următor.

Exemplul 10.3.5

Scorurile obţinute de subiecţi la itemi şi la testul compus din ei sunt notate în Tabelul 10.3.5.

Tabelul 10.3.5

Scorurile itemilor $i_1, i_2, \ldots, i_6$ pentru subiecţii $s_1, s_2, \ldots, s_{10}$

Item	Persoana										Dificultate
	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	
i_1	0	0	0	0	0	0	0	0	0	1	0.1
i_2	0	0	0	0	0	0	0	0	1	1	0.2
i_3	0	0	0	0	0	0	1	1	1	1	0.4
i_4	0	0	0	0	1	1	1	1	1	1	0.6
i_5	0	0	1	1	1	1	1	1	1	1	0.8
i_6	0	1	1	1	1	1	1	1	1	1	0.9
Scor total	0	1	2	2	3	3	4	4	5	6	

De data aceasta apar toate scorurile posibile ale testului, iar repartiţia lor de frecvenţe este cuprinsă între repartiţia uniformă şi cea normală.

CAPITOLUL 11
INFORMAŢII DESPRE TESTELE PSIHOLOGICE

11.1. MANUALELE TESTELOR PSIHOLOGICE

În mod obişnuit, orice test psihologic care este pus în circulaţie, pentru a fi utilizat în practică, este însoţit de două manuale:

- un *manual tehnic*, care conţine informaţii despre felul în care a fost construit testul, rezultatele analizei de itemi, valorile coeficienţilor de fidelitate şi concluziile studiilor de validitate;
- un *manual de utilizare*, care cuprinde instrucţiunile pentru administrarea testului şi indicaţii pentru interpretarea rezultatelor.

Standards for Educational and Psychological Tests din 1985 recomandă ca, pe lângă furnizarea acestor informaţii, manualele testelor psihologice să îndeplinească şi următoarele sarcini (Gregory, 1992):

a. să descrie motivele construirii testului şi să prezinte recomandări pentru utilizarea lui;

b. să atragă atenţia asupra eventualelor riscuri de folosire greşită a testului;

c. să citeze studii reprezentative privind utilizarea testului în situaţii obişnuite sau specifice;

d. să prezinte condiţiile (de studii, de experienţă etc.) pe care trebuie să le îndeplinească persoana care administrează testul şi cea care interpretează rezultatele acestuia;

e. să descrie relaţiile constatate între scorurile testului şi diverse criterii;

f. să prezinte posibilităţile de interpretare a rezultatelor testului cu ajutorul calculatorului.

11.2. INFORMAŢII NECESARE PENTRU ALEGEREA UNUI TEST PSIHOLOGIC

Atunci când sunt disponibile mai multe teste care servesc aceluiaşi scop, alegerea celui mai potrivit dintre ele impune luarea în considerare a unor criterii de ordin tehnic, practic şi economic. Faptul că un test este un instrument de măsură sau de predicţie bun, adecvat lotului care urmează să fie examinat şi

accesibil ca preţ, nu surmontează dificultăţile legate de administrarea şi de cotarea sa, ori de interpretarea rezultatelor.

Principalele aspecte care pot influenţa alegerea unui test psihologic sunt cuprinse în lista următoare (adaptată după Gronlund, 1967; Thorndike, Hagen, 1961):

A. *Date pentru identificarea testului*:

- denumirea testului;
- autorii testului (numele şi, eventual, autoritatea lor ştiinţifică);
- versiunea testului;
- data punerii în circulaţie;
- denumirea firmei care difuzează testul.

B. *Informaţii generale despre test*:

- funcţia testului: ce măsoară şi cum este definit constructul măsurat sau ce performanţe prezice;
- caracteristicile persoanelor care pot fi examinate cu acest test: zona geografică, vârsta, nivelul de şcolarizare etc.;
- timpul necesar pentru administrarea testului;
- modul de administrare: creion-hârtie, cu aparate, cu ajutorul calculatorului etc.;
- modul de calcul al scorurilor: manual, cu calculatorul etc.;
- cum se interpretează scorurile testului: obiectiv sau subiectiv;
- dacă există forme paralele ale testului, disponibile;
- costul achiziţionării testului;
- costul administrării testului pentru o persoană.

C. *Caracteristici tehnice*:

- calitatea itemilor testului: dacă folosesc o terminologie cunoscută celor care vor fi examinaţi cu acest test, dacă sunt lipsiţi de ambiguitate, dacă în itemii cu alegere multiplă răspunsurile greşite propuse nu sunt banale;
- validitatea: modul în care a fost studiată validitatea şi constatările făcute;
- fidelitatea: ce coeficienţi au fost calculaţi, în ce condiţii şi ce rezultate s-au obţinut;
- norme disponibile: ce tipuri de norme se utilizează, ce populaţii au fost folosite pentru etalonare, dacă există norme potrivite pentru populaţia căreia urmează să îi fie administrat testul.

D. *Caracteristici practice*:

- dificultăţi la administrare: dacă sunt necesare aparate sau condiţii speciale;

● dificultăţi la calculul scorurilor: dacă se aplică formule complicate şi nu există program pentru calculul automatizat al scorurilor sau dacă este un test subiectiv, care necesită o interpretare calitativă a răspunsurilor subiecţilor;
● dificultăţi la interpretarea scorurilor: dacă persoana care urmează să interpreteze scorurile trebuie să îndeplinească anumite condiţii de studii sau de experienţă;
● calitatea manualului testului: dacă există suficiente informaţii pentru utilizarea testului şi dacă acestea sunt clare.

E. *Alte consideraţii*:
● părerile unor persoane care au utilizat deja testul.

Majoritatea acestor informaţii se găsesc în manualele testului. Cele referitoare la existenţa formelor paralele ale testului sau a programelor pentru administrarea, cotarea şi/sau interpretarea rezultatelor testului se pot obţine din reviste de specialitate. Informaţiile privind distribuitorii testului, costul de achiziţie şi cel de utilizare (de administrare şi/sau de cotare) sunt specifice fiecărei ţări.

Foarte multe reviste publică articole despre validitatea şi fidelitatea testelor. Cel mai bun mijloc pentru identificarea lucrărilor referitoare la un anumit test îl constituie revista de rezumate *Psychological Abstracts*.

Se pot obţine multe informaţii despre testele psihologice din următoarele reviste:
● *Advances in Personality Assessment*;
● *The Journal of Psychoeducational Assessment*;
● *Psychology in the School*;
● *Educational and Psychological Measurement*;
● *Psychological Assessment: A Journal of Clinical and Consulting Psychology*;
● *The Journal of Clinical Psychology*;
● *The Journal of Personality Assessment*.

În *Revista de Psihologie* şi în *Revue Roumaine de Psychologie* sunt publicate articole despre studiile efectuate în România asupra unor teste psihologice originale sau traduse.

CAPITOLUL 12
CONSTRUIREA TESTELOR PSIHOLOGICE

12.1. ALGORITMUL GENERAL DE CONSTRUIRE A TESTELOR

12.1.1. Prezentarea paşilor algoritmului de construire a unui test

În general, construirea unui test psihologic sau a unei scale dintr-un test se desfăşoară în mai mulţi paşi, nesecvenţial, cu reveniri repetate până la realizarea calităţilor dorite (fidelitate, validitate, putere de discriminare, dificultate). Algoritmul general de construire a unui test este redat în figura 12.1.1.1.

Conţinutul tuturor paşilor algoritmului este determinat de direcţiile stabilite la primul pas, în special de scopul - teoretic şi practic - căruia trebuie să îi servească elaborarea testului. În funcţie de acesta se formulează itemii, se alege metoda de construire a testului şi se alcătuieşte o primă variantă a testului.

Gruparea de itemi formată se administrează unui eşantion de persoane extras din populaţia căreia îi este destinat testul. Se analizează răspunsurile obţinute la fiecare item, se identifică itemii necorespunzători (de exemplu, cei la care majoritatea persoanelor au răspuns identic sau, în cazul itemilor cu alegere multiplă, cei la care un răspuns greşit s-a dovedit a fi mai atractiv pentru subiecţi decât răspunsul corect) şi se caută cauza, pentru a vedea dacă este legată de item, de eşantionul de subiecţi sau de modul de administrare a itemilor. Analiza cauzelor erorilor poate scoate la iveală deficienţe ale unor itemi care în urma analizei statistice a rezultatelor păreau a fi bine construiţi (de exemplu, la un item s-a răspuns greşit pentru că un alt item conţine informaţii eronate, care i-au derutat pe subiecţi). Itemii fără greşeli se introduc în partea a doua a analizei de itemi, care aplică procedee specifice metodei alese pentru construirea testului (de exemplu, studiul validităţii itemilor sau calculul puterii de discriminare a acestora). Este posibil să fie puşi în evidenţă alţi itemi necorespunzători. Din itemii fără deficienţe se alcătuieşte un test, folosind un algoritm specific metodei de construcţie alese la primul pas al algoritmului.

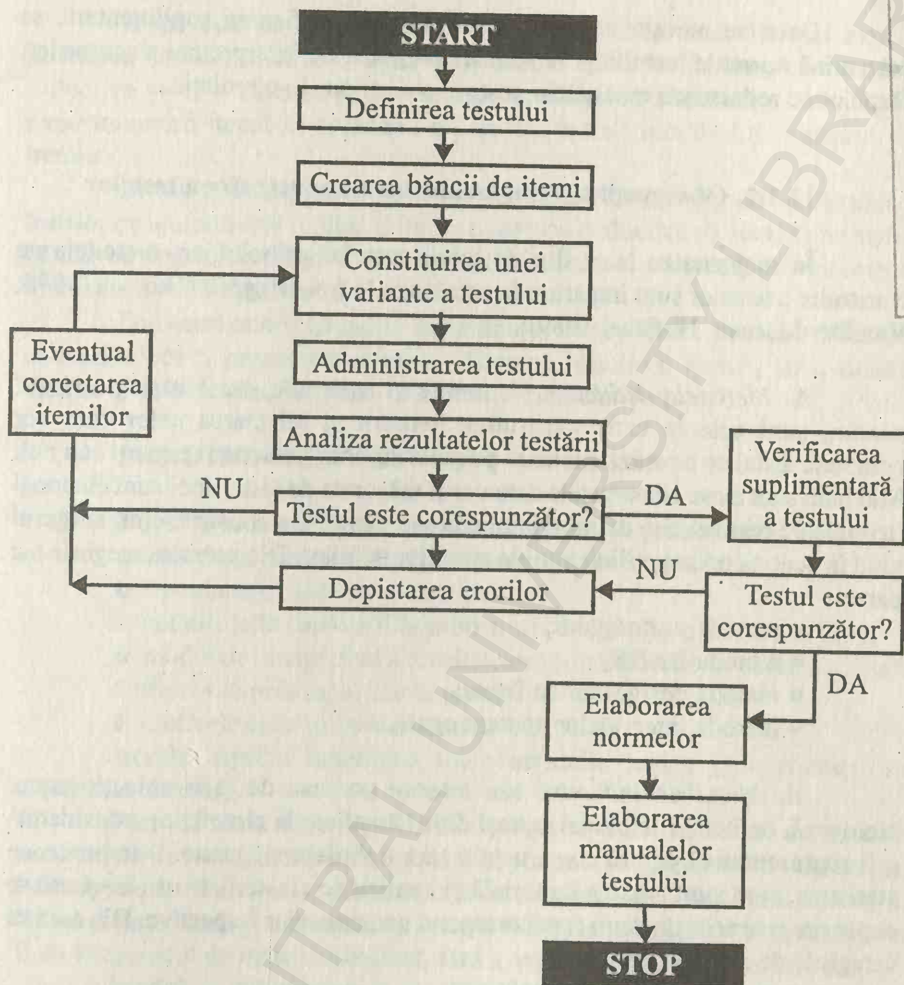


Figura 12.1.1.1
Algoritmul general de construire a unui test psihologic

Se studiază caracteristicile acestui test (fidelitatea, validitatea, puterea de discriminare, dificultatea) şi, dacă sunt considerate acceptabile, se administrează testul unui nou eşantion de persoane, similar primului, pentru a verifica dacă se păstrează calităţile.

Atunci când pentru unul din eşantioane testul apare ca fiind inacceptabil, se alcătuieşte o nouă variantă a testului şi se reia procedura descrisă mai sus.

Dacă se menţin calităţile testului şi la verificarea suplimentară, se determină normele testului şi se stabilesc regulile de interpretare a scorurilor. După ce se redactează manualele, testul poate fi pus în circulaţie.

12.1.2. Observaţii asupra metodelor de construire a testelor

În majoritatea lucrărilor de teoria testelor psihologice, metodele de construire a testelor sunt împărţite în următoarele trei categorii (Burisch, 1986; Van Der Maesen, Hofstee, 1989):

A. *Metodele deductive*, denumite şi *raţionale* sau *bazate pe simţul comun*, sunt cele în care construirea itemilor şi selectarea celor care vor compune testul se bazează exclusiv pe părerile unor persoane (experţi sau nu). Mai întâi sunt alese constructele care vor fi măsurate de teste, apoi sunt elaboraţi itemii, care realizează o definire implicită sau explicită a constructelor, singurul ghid în această operaţie fiind intuiţia autorilor itemilor. Din această categorie fac parte:

- metoda prototipului;
- metoda BARS;
- metoda designului cu faţete;
- metoda intervalelor aparent egale.

B. *Metodele inductive* sau *interne* pornesc de la o colecţie mare, eterogenă, de itemi (de obicei extraşi din teste aflate în circulaţie), considerată a fi reprezentativă pentru constructele care trebuiesc măsurate. Prin procedee statistice, cum sunt analiza factorială şi analiza de clusteri, se urmăreşte să se depisteze grupările de itemi care corespund constructelor respective. Din această categorie de metode fac parte:

- metoda consistenţei interne;
- metoda teoretică;
- metoda analizei criteriului intern.

C. *Metodele externe* sau *empirice* necesită precizarea, de la început, a relaţiilor care vor trebui să existe între scorurile testului care va fi construit şi diverse criterii externe. Se porneşte de la o colecţie mare, eterogenă, de itemi, se compară scorurile acestora cu criteriile şi se reţin cei care au comportarea pe care trebuie să o aibă întregul test. Din această categorie de metode fac parte:

- metoda criteriului extern;
- metoda analizei constructului.

După cum se observă, metodele deductive nu necesită nici experimentarea itemilor, nici efectuarea unor analize statistice, în timp ce metodele inductive şi cele externe se bazează pe prelucrarea statistică a rezultatelor experimentării itemilor, acordând o pondere redusă modului de construire a itemilor.

În multe situaţii este posibilă - şi chiar recomandabilă - construirea testelor cu ajutorul mai multor metode: o metodă deductivă va servi la formarea băncii de itemi, iar printr-o metodă internă sau/şi una externă se vor selecta itemii care vor alcătui testul.

Din acest motiv, în cadrul algoritmului de construire a testelor metodele deductive vor fi prezentate la pasul "Crearea băncii de itemi", iar metodele interne şi cele externe, la pasul "Analiza rezultatelor testului".

12.2. DEFINIREA TESTULUI

După ce se ia decizia de a construi un test, se precizează în detaliu:

a. funcţia pe care o va îndeplini (măsurare sau predicţie);
b. populaţia căreia îi este destinat;
c. condiţiile în care va fi administrat: când, unde, cum, în cât timp;
d. modul de interpretare a rezultatelor (prin raportare la obiective fixate a priori sau prin raportare la norme);
e. motivele care au impus crearea lui.

Aceste aspecte determină toate acţiunile legate de elaborarea şi experimentarea sa.

Pornind de la funcţia pe care o va realiza testul se decide care dintre caracteristicile sale psihometrice este mai importantă (fidelitatea sau validitatea), întrucât se ştie că un test nu poate avea în acelaşi timp şi o validitate relativă la criteriu foarte bună, şi o consistenţă internă ridicată. De asemenea, un test poate fi un instrument de măsură excelent, fără a avea vreo valoare predictivă.

Modul de interpretare a rezultatelor unui subiect este dependent de scopul căruia îi va servi testul. Atunci când testul verifică dacă anumite deprinderi sau cunoştinţe ale unei persoane depăşesc un prag minim necesar, scorul individual va fi raportat la bareme fixate a priori. Dacă, însă, testul prezice comportamentul unei persoane în situaţii specifice sau determină cât de eficient este individul într-un anumit domeniu, comparativ cu alte persoane din populaţia căreia îi aparţine subiectul, atunci interpretarea scorului individual se face prin raportare la norme.

Tot la acest prim pas al algoritmului de construire a testului se decide:

a. ce fel de scală de măsură va utiliza testul (nominală, ordinală sau de interval);

b. ce tipuri de itemi se vor include în test (în funcţie de ceea ce va măsura sau va prezice testul şi de faptul că pentru cotarea răspunsurilor se va utiliza sau nu calculatorul);

c. cum vor fi ordonaţi itemii (după dificultate sau aleator);

d. ce lungime va avea testul (în funcţie de caracteristicile populaţiei căreia i se va administra testul şi de timpul de administrare);

e. cum va fi administrat testul: cu sau fără limită de timp

A. Atunci când noul test trebuie să măsoare un construct, este de dorit ca la baza sa să stea o teorie structurală asupra personalităţii, care să reflecte şi importanţa pentru viaţă a diferenţelor individuale în privinţa acestuia (Angleitner, John, Löhr, 1986). Definirea constructului care va fi măsurat de test va specifica formele observabile relevante de manifestare a constructului respectiv (comportamente, interese, dorinţe) şi, în plus, va prezenta relaţiile care, conform teoriei, se aşteaptă să existe între acesta şi alte constructe (atât relaţii convergente, cât şi discriminante). Astfel de relaţii se pot deduce uşor dacă pentru constructul măsurat de test se construieşte o reţea nomologică.

Exemplu (Gregory, 1992).

A.S. Kaufman şi N.L. Kaufman, când şi-au propus să construiască un test pentru măsurarea inteligenţei, au formulat condiţiile impuse acestuia astfel:

- Să măsoare inteligenţa, având la bază o teorie puternică şi rezultate ale cercetărilor.
- Să facă distincţie între cunoştinţele dobândite şi abilitatea de a rezolva probleme cu care subiectul nu este familiarizat.
- Să furnizeze scoruri care să poată fi utilizate pentru intervenţii educaţionale.
- Să includă sarcini noi.
- Să fie uşor de administrat.
- Să fie cotat obiectiv.
- Să fie sensibil la diverse caracteristici ale preşcolarilor, ale grupurilor minoritare şi ale copiilor cu necesităţi speciale.

B. Un test care urmează să fie folosit în selecţia profesională trebuie să poată prezice corect performanţa persoanelor într-o anumită muncă. Pentru asemenea teste este necesar să se realizeze, de la început:

a. analiza muncii respective: să se obţină informaţii despre sarcinile de muncă şi despre cerinţele postului;

b. descrierea muncii: să se prezinte în detaliu, pentru munca respectivă, sarcinile, responsabilităţile şi condiţiile de lucru;

c. specificarea muncii: să se stabilească ce caracteristici fizice şi psihice sunt necesare desfăşurării muncii respective, alături de care se vor evidenţia (Kline, 1993):

- deprinderile şi cunoştinţele care nu sunt absolut necesare, deşi simţul comun ar sugera că ele sunt importante pentru îndeplinirea muncii (de exemplu, cunoştinţele de anatomie nu le sunt absolut necesare psihiatrilor, ele intervenind extrem de rar în activitatea lor zilnică):
- caracteristicile necesare datorită contextului în care se desfăşoară munca respectivă. (De exemplu, un flautist profesionist trebuie să fie capabil să lucreze într-o orchestră mare, să fie familiarizat cu repertoriul ultimelor trei secole, să poată să cânte în concerte după un număr mic de repetiţii, să poată suporta efortul unor călătorii lungi şi să fie dispus să lipsească mult de acasă. Aceste cerinţe sunt diferite de cele pentru un flautist de jazz sau pentru un flautist într-o formaţie muzicală mică.)

C. În cazul testelor de cunoştinţe, definirea testului constă în descrierea domeniului de conţinut. Adesea, se construieşte o tabelă de specificaţii, care indică repartiţia itemilor testului în funcţie de obiectivele comportamentale urmărite de test şi de temele a căror cunoaştere se verifică. O asemenea tabelă este prezentată în tabelul 12.2.1.

Tabela de specificaţii serveşte ca bază pentru a decide ce fel de itemi se vor introduce în test, întrucât pentru fiecare celulă a tabelului sunt mai potrivite anumite tipuri de itemi (de exemplu, capacitatea de explicare a unor fapte poate fi verificată mai bine prin itemi de tip eseu decât prin itemi cu răspunsuri închise).

Apariţia unui test nou trebuie să răspundă unei necesităţi reale. Se recurge la procedura - costisitoare - de creare a unui test atunci când nu există un test care să îndeplinească scopul urmărit (măsurarea unui anumit construct, verificarea cunoştinţelor dintr-un anumit domeniu sau predicţia unui anumit comportament), dacă testele disponibile sunt dificil de administrat, de cotat sau/şi de interpretat ori sunt prea scumpe. Uneori se construiesc forme paralele ale unor teste existente, fie pentru a nu administra repetat acelaşi test, fie pentru a se evita transmiterea itemilor de la un eşantion de subiecţi la altul şi a putea păstra secret conţinutul testului (de exemplu, în cazul testelor de cunoştinţe utilizate la diverse concursuri).

Tabelul 12.2.1

Tabela de specificaţii pentru verificarea cunoştinţelor la tema "Teoria testelor psihologice"

Obiectivul comportamental	Subiectul		
	Validitatea testelor	Fidelitatea testelor	Analiza de itemi
Cunoaşterea terminologiei	Validitatea relativă la: -construct; -conţinut; -criteriu. (3 itemi)	Consistenţa internă. Stabilitatea în timp. Forme paralele ale testelor. (3 itemi)	Dificultatea unui item. Validitatea unui item. (2 itemi)
Cunoaşterea faptelor specifice	Metoda de studiere a validităţii: -predictive; -concurente. (2 itemi)	Metodele: -test-retest; -formelor paralele; -analiza consistenţei interne. (3 itemi)	Procedura de calcul pentru: -indicele de dificultate; -puterea de discriminare; -coeficientul de validitate. (3 itemi)
Înţelegere	Explicarea legăturii dintre validitate şi fidelitate. Explicarea efectelor restrângerii intervalului de valori asupra coeficientului de validitate. (2 itemi)	Explicarea utilizării coeficientului de fidelitate pentru interpretarea scorurilor la test. (1 item)	Explicarea deosebirii dintre puterea de discriminare şi validitatea itemilor. (1 item)
Aplicare	Calculul unui coeficient de validitate. (1 item)	Aplicarea formulei lui Spearman-Brown. Determinarea numărului de itemi care trebuiesc adăugaţi unui test pentru dublarea valorii coeficientului de fidelitate. (2 itemi)	Calulul indicelui de dificultate şi al puterii de discriminare. (2 itemi)
TOTAL	8 itemi	9 itemi	8 itemi

12.3. CREAREA BĂNCII DE ITEMI

Itemii care se introduc în banca de itemi pot să provină din mai multe surse:

a. sunt formulaţi de psihologi, pe baza unei teorii (dacă testul măsoară un construct psihologic);

b. sunt elaboraţi de specialişti din domeniul în care va fi aplicat testul (dacă acesta evaluează deprinderi ori cunoştinţe însuşite în urma unui program de instruire sau dacă se utilizează în selecţia profesională);

c. sunt creaţi de diverse persoane (nu neapărat experţi), ca ilustrări ale comportamentelor pe care acestea le consideră caracteristice unui anumit construct;

d. sunt extraşi din teste existente, care au fost administrate unui lot de persoane asemănător celui căruia îi este destinat noul test.

Ultimul procedeu se aplică de obicei atunci când s-a constatat că itemii respectivi conţin un factor comun care corespunde constructului pe care îl va măsura testul sau când aceştia îndeplinesc cerinţele care decurg din scopul în care va fi utilizat testul. De exemplu:

- au cu alte variabile legături conforme teoriei care stă la baza constructului

sau

- se încadrează în domeniul de conţinut al testului

sau

- corelează semnificativ cu variabila criteriu

sau

- discriminează între două grupe contrastante de persoane.

Înainte de introducerea itemilor în bancă se stabileşte modul în care vor fi cotaţi. Uneori, se apreciază dificultatea acestora, atunci când ea este importantă pentru alegerea cotelor care vor fi acordate răspunsurilor sau dacă în testul care va fi format itemii vor fi aranjaţi în ordinea crescătoare a dificultăţii.

Se recomandă ca numărul itemilor introduşi în bancă să fie cel puţin cu 20% mai mare decât lungimea testului (Aiken, 1994).

După ce itemii au fost scrişi, este bine ca ei să fie revăzuţi şi editaţi de o persoană care nu a participat la elaborarea lor. Ea poate sesiza eventualele erori şi poate face propuneri pentru îmbunătăţirea itemilor.

Există firme specializate în crearea de teste, cum este, de exemplu, Educational Testing Service. Acestea folosesc pentru scrierea itemilor persoane care au cunoştinte din domeniul la care se va referi testul şi, în plus, au deprinderea de a crea itemi.

Cei care doresc să elaboreze itemi de test pot utiliza drept model itemii din testele standardizate comercializate.

Prezentăm în continuare câteva metode de creare a băncii de itemi, din cadrul metodelor deductive de construire a testelor. În cazul metodelor interne şi externe de construire a testelor, alcătuirea băncii de itemi este mai simplă: itemii -noi sau extraşi din alte teste- nu mai sunt verificaţi înainte de a fi introduşi în bancă, întrucât după administrarea lor unui eşantion de subiecţi rezultatele obţinute vor fi prelucrate statistic.

Metoda prototipului (Oosterveld, Vorst, 1995) se utilizează pentru construirea chestionarelor de personalitate.

Mai întâi sunt generaţi itemii. În acest scop, persoanelor care participă la această acţiune li se cere să descrie comportamente tipice pentru indivizi care au valori foarte mari la constructul măsurat de test. De exemplu, pentru elaborarea unui test care să evalueze dominanţa, se poate folosi instrucţiunea: "Gândiţi-vă la trei bărbaţi (femei) care sunt cei mai dominanţi dintre cunoştinţele dumneavoastră. Descrieţi cinci acţiuni sau comportamente ale acestora prin care puteţi exemplifica de ce îi consideraţi dominanţi."

Se analizează apoi mulţimea de descrieri obţinute, se elimină redundanţele şi se corectează greşelile de exprimare. Fiecare descriere rămasă va constitui un item.

Pasul următor constă din evaluarea itemilor de către mai multe persoane. Se acordă fiecărui item câte o cotă, de exemplu între 1 şi 5, care să reflecte cât de tipic este comportamentul descris de item pentru cei care au valori mari la constructul măsurat de test.

Se calculează pentru fiecare item media cotelor primite de la toţi evaluatorii şi se reţin în bancă itemii care au mediile cele mai mari.

O variantă a acestei metode (Angleitner, John, Löhr, 1986) porneşte de la definirea mai multor constructe, pe baza unui model structural al comportamentului interpersonal. Un grup numeros de subiecţi elaborează, pentru fiecare construct, itemi care descriu manifestări comportamentale specifice constructului. Apoi, fiecare item este analizat de persoane dintr-un alt grup şi este atribuit constructului pe care se consideră că îl ilustrează cel mai bine. Sunt reţinuţi în banca de itemi doar acei itemi care au fost alocaţi, cu frecvenţă mare, constructului pentru ilustrarea căruia au fost creaţi.

Metoda BARS (Behaviorall Anchored Rating Scales) a fost aplicată la început pentru construirea unor scale de apreciere a comportamentului legat

de muncă. Fiecare scală corespundea unei dimensiuni a activităţii profesionale, iar ancorele erau exemple de niveluri de performanţă. Mai nou, această metodă este utilizată şi pentru elaborarea unor chestionare de autoevaluare comportamentală bazate pe ancore. Este cazul unor chestionare care măsoară constructe multidimensionale. În cadrul lor, fiecare scală se referă la câte o dimensiune.

Algoritmul de construire a unui asemenea chestionar este compus din următorii paşi (Landy, Rastegary ş.a., 1991):

(1) Un grup de specialişti în domeniul pentru care se construieşte chestionarul încearcă să identifice dimensiunile constructului. Fiecare dimensiune primeşte un nume, fără a fi, însă, definită.

(2) Lista cu denumiri este revăzută. Se elimină redundanţele. Unele denumiri sunt combinate între ele şi reformulate, pentru a se obţine categorii mai uşor de înţeles.

(3) Pentru fiecare "etichetă" a constructului se generează o definiţie. Se contopesc dimensiunile pentru care, în urma analizei definiţiilor, se constată existenţa unor suprapuneri.

(4) Se generează exemple pentru diverse niveluri de manifestare a dimensiunilor. De pildă, se pot fixa trei niveluri (înalt, mediu şi scăzut) şi i se cere fiecărei persoane din grupul de specialişti să formuleze câte trei ancore pentru fiecare nivel al fiecărei dimensiuni.

(5) Se elimină exemplele duble.

(6) Toate ancorele rămase sunt prezentate unui grup de evaluatori care au sarcina de a identifica dimensiunea la care se referă fiecare. Sunt reţinute exemplele pentru care cel puţin 75% dintre persoane au recunoscut dimensiunea măsurată.

(7) Se elimină dimensiunile cărora li s-au repartizat doar puţine ancore.

(8) Un alt grup de persoane atribuie valori dintr-o mulţime fixată (de exemplu, $\{1,2,...,7\}$) fiecărei ancore. Ponderea acordată trebuie să reflecte nivelul de dezvoltare al constructului măsurat de chestionar pentru care exemplul respectiv este potrivit.

(9) Pentru fiecare ancoră se calculează media şi abaterea standard a cotelor primite. Se elimină exemplele care au abateri standard mari (de obicei, cele pentru care $\sigma>1$); acestea au fost interpretate diferit de către evaluatori.

(10) Se ordonează ancorele fiecărei dimensiuni crescător după medie. Se formează scalele -câte una pentru fiecare dimensiune-reţinând între 4 şi 8 ancore, repartizate între nivelul cel mai scăzut şi nivelul cel mai ridicat de manifestare a dimensiunii respective.

Asemănătoare cu metoda BARS este **metoda ordonării comportamentelor de către experţi** (Gregory, 1992). Aceasta serveşte la crearea unor

teste în care itemii sunt aşezaţi în ordinea crescătoare a dificultăţii şi care au proprietatea că ori de câte ori un item este potrivit pentru o persoană, sunt potriviţi într-o oarecare măsură şi toţi itemii care îl preced. Scorul unui subiect este dat de numărul de ordine al celui mai dificil item care i se potriveşte.

Un exemplu de astfel de test îl reprezintă "Scala celui mai bun răspuns verbal" din Glasgow Coma Scale for Recording Depth of Coma, un test utilizat în spitale pentru evaluarea gradului de afectare a creierului în urma unui traumatism. Itemii componenţi sunt:

1 - nici un răspuns;
2 - răspuns care nu poate fi înţeles;
3 - răspuns nepotrivit;
4 - răspuns confuz;
5 - răspuns la obiect.

Scorul scalei este stabilit de medic în urma observării pacientului.

Metoda designului cu faţete (Oosterveld, Vorst, 1995) are ca punct de pornire analiza constructului pe care îl va măsura testul şi deducerea aspectelor mai importante ale acestuia, denumite "faţete". De exemplu, faţetele pot fi "situaţia" şi "răspunsul".

Pentru fiecare faţetă se identifică elementele cele mai importante. Faţeta "răspuns", de pildă, poate avea ca elemente: "cognitiv", "psihologic", "comportamental" şi "afectiv".

Se construieşte apoi o matrice, cu atâtea dimensiuni câte faţete au fost alese, în care fiecare celulă corespunde unui element din fiecare faţetă. De exemplu, o celulă poate corespunde răspunsului "cognitiv" şi situaţiei "îmbolnăvire gravă". Această matrice este denumită "designul cu faţete".

Se scriu unul sau mai mulţi itemi pentru fiecare celulă a matricii.

Metoda intervalelor aparent egale (Gregory, 1992; Kline, 1993) a fost propusă de L.L Thurstone pentru construirea scalelor de evaluare a atitudinilor.

Se începe cu formarea unei colecţii, cât mai bogate, de itemi dihotomici, care se referă, toţi, la atitudinea faţă de un acelaşi obiect (persoană, instituţie, situaţie sau noţiune abstractă). La fiecare item se poate răspunde doar prin "adevărat" sau "fals". Răspunsul ales de subiect este privit ca o reflectare a atitudinii sale, pozitive sau negative. De exemplu, pentru cercetarea atitudinii faţă de filmele muzicale se pot folosi itemii: "Un film muzical mă face optimist." şi "Este o pierdere de vreme să urmăresc un film muzical".

Un grup de experţi analizează fiecare item şi apreciază cât de favorabilă este atitudinea studiată la persoanele care răspund cu "adevărat", acordând o

cotă cuprinsă între 1 (extrem de nefavorabilă) şi 11 (extrem de favorabilă). Evaluatorilor li se cere să privească cele 11 cote posibile ca fiind echidistante.

Pentru fiecare item se calculează media şi dispersia cotelor primite. Deoarece o dispersie mare indică o neconcordanţă a părerilor, itemii cu dispersii mari sunt eliminaţi. Media cotelor va fi considerată drept scor corespunzător răspunsului "adevărat".

Fiecărei persoane examinate cu un astfel de test i se cere să indice itemii la care răspunsul său este "adevărat". Scorul total al testului pentru un subiect va fi egal cu media scorurilor itemilor indicaţi.

Scalele de acest tip se compun din 20-30 de itemi, aleşi astfel încât scorurile corespunzătoare răspunsului "adevărat" să fie aproximativ echidistante în intervalul [1, 11].

O metodă asemănătoare celei descrise mai sus utilizează mediana în locul mediei, atât pentru determinarea scorurilor itemilor, cât şi pentru calculul scorului unui subiect.

12.4. CONSTRUIREA UNEI VARIANTE A TESTULUI

Înainte de selectarea itemilor din banca de itemi, pentru alcătuirea unei variante a testului, se stabileşte lungimea acestuia. Numărul de itemi care îl vor compune are influenţă atât asupra timpului de rezolvare a testului, cât şi asupra calităţilor sale psihometrice (fidelitate, putere de discriminare).

Lungimea testului este aleasă în funcţie de timpul disponibil pentru administrarea testului, care este fixat, în general, ţinând cont de condiţiile în care va fi utilizat testul (într-o oră de clasă, la un concurs de admitere într-o instituţie de învăţământ etc.).

Timpul de rezolvare a unui test este dependent nu numai de numărul itemilor componenţi, ci şi de:

- vârsta subiecţilor;
- deprinderea subiecţilor de a citi;
- lungimea şi dificultatea itemilor;
- modul de formulare a răspunsurilor şi lungimea acestora.

În practică s-a constatat că un test de dificultate medie, administrat elevilor din gimnaziu, necesită cam un minut pentru rezolvarea unui item cu alegere multiplă sau cu răspuns lacunar şi aproximativ 30 de secunde pentru alegerea răspunsului la un item dihotomic. Prin urmare, într-o oră de clasă de 50 de minute se pot administra teste compuse din 100 de itemi dihotomici sau 50 de itemi cu alegere multiplă ori cu răspuns lacunar. În acelaşi interval de

timp se pot rezolva 5-6 itemi cu răspunsuri deschise, care necesită, pentru fiecare răspuns, cam jumătate de pagină (Aiken, 1994).

În orice caz, o testare nu are voie să depăşească o oră pentru elevii din şcoala primară şi o oră şi jumătate pentru cei din gimnaziu şi din liceu. Întrucât copiii preşcolari şi elevii din clasele primare nu pot să rămână atenţi la sarcinile testului mai mult de 30 de minute, dacă rezolvarea testului necesită un timp mai lung, se va recurge la administrarea testului în mai multe sesiuni de lucru (Aiken, 1994).

Există diverse posibilităţi pentru construirea testului. De exemplu:

- se extrag aleator itemii din bancă

sau

- dacă a fost întocmită o tabelă de specificaţii, se respectă frecvenţele înscrise în fiecare celulă a sa

sau

- dacă dificultatea itemilor a fost apreciată înainte de introducerea lor în banca de itemi, se aleg itemii astfel încât valorile dificultăţilor lor să fie repartizate cât mai uniform pe intervalul valorilor posibile sau dorite (uneori se urmăreşte alcătuirea unui test dificil sau a unuia uşor).

Atunci când testul format conţine itemi de mai multe tipuri, este recomandabil ca toţi itemii de acelaşi tip să fie grupaţi, pentru ca subiectul să nu mai facă şi efortul de a identifica, la fiecare item, modul în care trebuie să răspundă. Acest procedeu facilitează, de asemenea, administrarea şi cotarea testului.

Aranjarea itemilor în ordinea crescătoare a dificultăţii îi ajută pe unuii subiecţi, crescându-le motivaţia pentru rezolvarea testului: văzând că pot rezolva primii itemi, ei vor anticipa obţinerea succesului în continuare şi vor avea curajul de a aborda şi itemii mai dificili. La testele administrate cu limită de timp şi la cele foarte dificile, acest mod de aranjare are ca efect creşterea scorurilor testului. Dar, se pare că efectul său este scăzut sau nul în cazul itemilor cu alegere multiplă.

Se va decide apoi cum vor fi aranjaţi itemii în pagină, pentru a putea fi uşor de citit, dar şi pentru a permite utilizarea economică a hârtiei. În funcţie de locul unde vor fi marcate răspunsurile (pe broşura testului sau pe foaia de răspuns) se va stabili aşezarea în pagină a poziţiilor care vor conţine răspunsurile, astfel încât corectarea să se facă uşor, eventual cu ajutorul unui şablon sau al calculatorului.

După alcătuirea testului, se vor redacta instrucţiunile de administrare. Acestea trebuie să fie clare şi să furnizeze suficiente detalii, pentru a nu fi necesare comentarii suplimentare în momentul testării.

Instrucţiunile de lucru vor conţine informaţii privind scopul testării, modul de lucru, din câţi itemi se compune testul, cât timp este alocat pentru răs-

puns şi cum se cotează răspunsurile. De asemenea, se va menţiona dacă este sau nu recomandabil să se ghicească răspunsul, atunci când nu se cunoaşte răspunsul corect sau cel mai potrivit, şi dacă este sau nu permis să se omită răspunsul.

Autorul testului are sarcina de a utiliza cele mai potrivite elemente grafice (sublinieri, caractere mai mari, culori contrastante) pentru a atrage atenţia asupra anumitor părţi din instrucţiunile de administrare, aşa încât cei care le vor citi subiecţilor să ştie cum să îşi modifice intonaţia.

12.5. ADMINISTRAREA TESTULUI

Administrarea testului se va face în aceleaşi condiţii în care va fi utilizat în practică. Se obişnuieşte ca broşura testului şi foaia de răspuns să le fie înmânate subiecţilor doar după citirea instrucţiunilor de administrare a testului şi rezolvarea eventualelor nelămuriri. Mai întâi se dă foaia de răspuns şi se repetă instrucţiunile privind modul de completare a răspunsurilor. Apoi, fiecare subiect va primi broşura testului, aşezată cu faţa în jos. Ea va fi întoarsă doar la comandă, de toate persoanele în acelaşi timp.

Este foarte important să li se spună subiecţilor ce informaţii de identificare trebuie să completeze (nume, sex, vârstă, funcţie etc.) şi unde.

În general (nu numai în faza de construire a testului), dacă testele se administrează în clasă, elevii trebuie să fie informaţi din timp când şi unde se va face testarea, ce fel de test se va administra (oral, scris etc.) şi la ce se referă acesta, pentru a se putea pregăti intelectual, emoţional şi fizic.

12.6. ANALIZA REZULTATELOR TESTĂRII

Se face o analiză atentă a modului în care a decurs testarea, urmărindu-se dacă instrucţiunile de administrare au fost complete şi dacă timpul de lucru a fost suficient. Se prelucrează apoi răspunsurile primite de la subiecţi şi se încercă să se construiască un test, pe baza celui experimentat. Se procedează astfel:

(1) Se face analiza de itemi.
(2) Se elimină itemii necorespunzători.
(3) Eventual, dintre itemii rămaşi se selectează cei care satisfac anumite condiţii, prin care se asigură o valoare cât mai bună a unui anumit parametru al testului.

(4) Se analizează apoi caracteristicile psihometrice (fidelitatea, validitatea, dificultatea, puterea de discriminare) ale testului obţinut.

12.6.1. Analiza de itemi

Analiza de itemi are ca obiective esenţiale descifrarea mecanismelor cognitive aplicate de subiecţi pentru formularea răspunsurilor la itemi şi verificarea calităţilor itemilor ca instrumente de măsură sau de predicţie. Ea furnizează informaţiile pe baza cărora se selectează itemii care vor intra în componenţa testului.

Este important de înţeles rolul suplimentar al analizei de itemi în cazul testelor care măsoară constructe. Dacă la început se dispune de o descriere, mai mult sau mai puţin detaliată a constructului, care serveşte ca bază pentru scrierea itemilor, după analizarea itemilor se fac inferenţe asupra constructului pornind de la relaţiile constatate între răspunsurile la itemi. Deci, deşi construirea testului începe cu definirea teoretică a constructului, constructul ajunge să fie o inferenţă fundamentată pe consistenţa răspunsurilor la itemi (Steinberg, Thissen, 1996).

Alegerea procedurilor statistice aplicate în cadrul analizei de itemi şi interpretarea rezultatelor obţinute sunt determinate de tipul itemilor, de funcţia pe care o va îndeplini testul şi de calităţile psihometrice pe care trebuie să le aibă acesta.

Pentru fiecare item, se efectuează una sau mai multe dintre următoarele prelucrări:

a. calculul frecvenţei cu care au fost alese răspunsurile şi identificarea cauzelor care au determinat frecvenţe foarte mari sau foarte mici;
b. inspectarea textului itemului pentru depistarea termenilor ambigui, a absenţei unor informaţii necesare pentru formularea unui răspuns corect şi identificarea răspunsurilor propuse neplauzibile (pe care nu le alege nimeni) sau evidente (pe care le aleg toţi cei testaţi);
c. verificarea comportării itemului faţă de diverse criterii;
d. determinarea capacităţii itemului de a face deosebire între examinaţii care au nivel scăzut şi cei care au nivel ridicat la variabila măsurată sau prezisă de test.

Vom descrie în continuare procedeele statistice ale analizei de itemi. Pentru simplificarea expunerii, vom desemna prin expresia "rezolvare corectă/greşită a itemului" indicarea de către subiect a acelui răspuns care corespunde nivelului înalt/scăzut al variabilei măsurate sau prezise de test.

12.6.1.1. Indicele de dificultate al unui item

Cea mai obişnuită măsură a dificultăţii itemului este frecvenţa relativă a persoanelor care au rezolvat corect itemul, din totalul celor cărora li s-a administrat testul (Murphy, Davidshofer, 1987).

Indicele de dificultate astfel definit se notează cu *p* şi reprezintă scorul mediu care s-ar obţine la item dacă s-ar acorda un punct pentru răspunsul corect şi zero puncte în cazul răspunsului greşit sau omis. Se observă că *p* este cuprins întotdeauna între 0 şi 1.

Davis (1955) recomandă ca, atunci când testul este administrat cu limită de timp şi există persoane care nu au reuşit să îl parcurgă integral, la calculul frecvenţei relative *p* pentru un anumit item să se utilizeze drept numitor numărul examinaţilor care au citit întrebarea şi nu efectivul lotului căruia i s-a administrat testul. Pentru fiecare persoană vor fi considerate necitite toate întrebările aflate după ultimul item rezolvat.

Indicele de dificultate este o caracteristică atât a itemului, cât şi a populaţiei căreia i s-a administrat testul. Un item cu răspunsuri la alegere care cere rezolvarea unei ecuaţii de gradul al doilea va fi rezolvat corect de majoritatea elevilor din clasa a X-a, în timp ce elevii din clasa a III-a vor omite răspunsul sau vor alege unul greşit, deci indicele său de dificultate va fi diferit de la o populaţie la alta.

Valorile extreme ale lui *p* restrâng variabilitatea scorurilor la test. Itemii cu $p=0$ (care nu au fost rezolvaţi corect de nici o persoană) şi cei cu $p=1$ (care au fost rezolvaţi corect de toţi cei examinaţi) nu contribuie la măsurarea variabilei vizate de test şi, prin urmare, sunt inutili.

Dispersia scorurilor la un test format din itemi necorelaţi liniar între ei este maximă atunci când toţi itemii au indicele de dificultate egal cu 0.5.

În general, itemii al căror indice de dificultate ia valori în intervalul [0.3, 0.7] permit o diferenţiere bună între subiecţi (Gregory, 1992).

Pentru itemii dihotomici şi pentru cei cu alegere multiplă care au un singur răspuns corect se recomandă să se păstreze în test itemii al căror indice de dificultate este apropiat de valoarea $0.5+g/2$, unde *g* reprezintă probabilitatea de a ghici răspunsul corect, adică este egal cu raportul 1/ numărul de răspunsuri posibile (Gregory, 1992).

Atunci când testul va fi utilizat pentru selecţia persoanelor care vor face parte dintr-un grup extrem (cu valori foarte mari sau foarte mici ale caracteristicii măsurate de test), se va forma testul din itemi care au indicele de dificultate mai mic decât 0.3 sau mai mare decât 0.7.

Pentru întrebările cu alegere multiplă din testele de aptitudini sau de cunoştinţe, care pretind selectarea unui singur răspuns, dacă se acceptă

presupunerea că examinaţii care nu cunosc răspunsul corect aleg la întâmplare dintre cele n răspunsuri propuse, se poate calcula indicele de dificultate cu următoarea formulă:

$$p_c = \frac{np-1}{n-1} \qquad (12.6.1.1.1)$$

unde

p este frecvenţa relativă a persoanelor care au răspuns corect la item.

p_c reprezintă scorul mediu care s-ar obţine la item dacă s-ar acorda un punct pentru răspunsul corect şi $-1/(n-1)$ puncte pentru răspunsurile greşite sau omise.

Guilford (1965) propune să se utilizeze această formulă atunci când trebuie să se compare, în ceea ce priveşte dificultatea, itemi care au numere diferite de răspunsuri propuse. Intuitiv, dacă doi itemi se referă la acelaşi lucru, sunt bine formulaţi, aşa încât răspunsurile propuse sunt la fel de atractive pentru cei care nu cunosc răspunsul corect, dar unul propune n_1 răspunsuri, iar celălalt n_2 răspunsuri, $n_1 < n_2$, ar fi de aşteptat ca primul item să fie rezolvat corect de mai mulţi subiecţi decât al doilea. Dacă, totuşi, frecvenţa relativă a rezolvărilor corecte este aceeaşi pentru ambii itemi, înseamnă că primul item este mai dificil. Acest lucru este pus în evidenţă de indicele p_c, după cum rezultă din inegalitatea:

$$\frac{n_1 p-1}{n_1-1} \le \frac{n_2 p-1}{n_2-1} \qquad (12.6.1.1.2)$$

Indicii de dificultate definiţi până acum nu sunt potriviţi în cazul întrebărilor care cer mai multe răspunsuri, întrucât nu iau în considerare decât trei situaţii: item rezolvat integral corect, item greşit şi item omis. Rezolvările parţial corecte sunt considerate greşite, indiferent câte răspunsuri corecte s-au dat din cele k cerute. Pentru a evita acest inconvenient şi a putea compara asemenea itemi, se utilizează următoarea formulă de calcul al indicelui de dificultate:

$$p_m = \frac{1}{k}\sum_{i=1}^{k} i p_i \qquad (12.6.1.1.3)$$

unde:

p_i = frecvenţa relativă a celor care au dat i răspunsuri corecte.

p_m reprezintă scorul mediu al itemului dacă s-ar acorda $1/k$ puncte pentru fiecare răspuns corect ales.

Pentru k=1, p_m coincide cu p, frecvenţa relativă a celor care au rezolvat corect itemul.

p_m astfel definit variază între 0 (când nu a fost indicat nici măcar un răspuns corect) şi 1 (când toate persoanele au indicat cele k răspunsuri corecte). p_m are valoare mare atunci când frecvenţa relativă a celor care au dat toate răspunsurile corecte este mare.

În cazul itemilor cu alegere multiplă este bine să se calculeze şi frecvenţa relativă cu care a fost ales fiecare răspuns propus, pentru a identifica răspunsurile corecte care au fost alese de foarte puţini subiecţi şi răspunsurile greşite alese de multe persoane.

Există situaţii, cum este cazul scalelor Likert, în care itemul pretinde alegerea unui singur răspuns, dar răspunsurile propuse nu sunt dihotomizate în "corect"/ "greşit" sau "adevărat"/ "fals", ci fiecare răspuns primeşte un alt punctaj. Atunci, în locul indicelui de dificultate al itemului se vor calcula scorul mediu şi abaterea standard a scorurilor itemului. În mod obligatoriu se va analiza şi forma repartiţiei scorurilor itemului, pentru a depista cazurile când sunt alese cu frecvenţe mari două sau mai multe răspunsuri care ar trebui să exprime lucruri foarte deosebite între ele şi a determina cauza.

O imagine mai clară asupra modului în care persoanele examinate au rezolvat itemul o oferă energia informaţională a itemului, calculată cu formula (Cresin, 1966):

$$E = \frac{\sum_{i=1}^{k}\left(\frac{p_i}{1-o}\right)^2 - \frac{1}{k}}{1-\frac{1}{k}} \qquad (12.6.1.1.4)$$

unde:

p_i = frecvenţa relativă a celor care au ales răspunsul i, i=1,2,...,k.

o = frecvenţa relativă a celor care nu au răspuns la item.

Energia informaţională variază între 0 (când toate persoanele au răspuns la item şi toate frecvenţele relative p_i sunt egale) şi 1 (când toate persoanele au ales acelaşi răspuns). Ea este mică atunci când toate răspunsurile propuse au fost alese cu frecvenţe apropiate şi este mare atunci când persoanele au ales un număr mic de răspunsuri.

12.6.1.2. Analiza răspunsurilor incorecte

La o întrebare cu răspunsuri la alegere dintr-un test de aptitudini sau de cunoştinţe este important să se verifice dacă factorii de distragere (răspunsurile incorecte din lista de răspunsuri propuse) îşi îndeplinesc într-adevăr misiunea.

Un procedeu de lucru recomandat în cazul întrebărilor care cer selectarea unui singur răspuns din *n* răspunsuri propuse este următorul (de Landsheere, 1975; Murphy, Davidshofer, 1987):

a. Pentru fiecare răspuns greşit *i*, se calculează:

• frecvenţa observată, $f_{o,i}$, egală cu frecvenţa relativă a cazurilor în care a fost ales:

$$f_{o,i} = \frac{N_i}{N} \qquad (12.6.1.2.1)$$

şi

• frecvenţa teoretică, f_t, egală cu frecvenţa relativă a cazurilor în care ar fi fost ales dacă toţi cei care au dat un răspuns greşit ar fi ales răspunsul greşit la întâmplare:

$$f_t = \frac{N-N_C}{N(n-1)} \qquad (12.6.1.2.2)$$

unde:

N = numărul de subiecţi care au răspuns la item;
N_C = numărul de subiecţi care au ales răspunsul corect;
N_i = numărul de subiecţi care au ales răspunsul *i*.

Toate răspunsurile greşite ale itemului au aceeaşi frecvenţă teoretică.

b. Se apreciază că întrebarea este bine construită dacă pentru fiecare răspuns greşit frecvenţa observată $f_{o,i}$ este aproximativ egală cu frecvenţa teoretică f_t.

Exemplu

Tabelul următor prezintă frecvenţele de răspuns la un item care propune 4 răspunsuri:

Răspuns		Frecvenţa alegerii fiecărui răspuns	Frecvenţa observată a fiecărui răspuns
	1	58	0.29
	2	4	0.02
corect	3	108	-
	4	30	0.15

Frecvenţa teoretică a fiecărui răspuns greşit este egală cu

$$f_t = \frac{200-108}{200*3} = 0.153$$

Răspunsul greşit 4 are frecvenţa observată apropiată de frecvenţa teoretică. În schimb, răspunsurile 1 şi 2 au fost alese de mult mai mulţi, respectiv mult mai puţini subiecţi decât era de aşteptat, dacă s-ar fi ghicit răspunsul. Ar trebui analizate răspunsurile greşite pentru a se depista cauzele abaterilor frecvenţelor observate de la frecvenţa teoretică.

Se consideră că un răspuns greşit pe care nu îl aleg decât foarte puţine persoane nu îşi îndeplineşte rolul de factor de distragere întrucât este identificat cu uşurinţă drept greşit şi, în acest fel, scade dificultatea sarcinii subiectului, care nu mai are de ales răspunsul corect dintre cele *n* răspunsuri propuse, ci dintre *n*-1. Un asemenea răspuns trebuie înlocuit.

Un răspuns greşit ales de un număr mare de persoane poate conţine o formulare incorectă sau incompletă, care îl face să semene cu răspunsul corect, dar, în acelaşi timp, poate semnala existenţa unei erori la răspunsul corect, care îi determină pe subiecţi să îl respingă şi să aleagă unul dintre celelalte răspunsuri.

Există şi situaţii în care unele răspunsuri greşite sunt formulate intenţionat astfel încât să se deosebească de răspunsul corect doar prin câte un element subtil. Ele vor fi alese frecvent, dar întrebările de acest tip sunt utile pentru a-i identifica pe subiecţii cu un nivel înalt al variabilei măsurate de test (de exemplu pentru selecţia elevilor în lotul olimpic).

La baza algoritmului descris mai sus a stat următorul raţionament: Subiecţii cărora li se prezintă o întrebare cu alegere multiplă se împart în două categorii; unii cunosc răspunsul corect şi îl marchează, ceilalţi aleg la întâmplare din lista de răspunsuri propuse. În realitate, însă, modul de selecţie a răspunsului diferă de la un subiect la altul. Unele persoane aleg un răspuns eronat cu

convingerea că acesta este cel corect. Alţii sunt capabili să identifice unul sau mai multe răspunsuri greşite şi să le elimine, alegând la întâmplare dintre celelalte. În sfârşit, mai sunt şi persoane care aleg răspunsul bazându-se pe alte considerente, cum ar fi lungimea răspunsurilor propuse, precizia limbajului folosit în formularea acestora, poziţia răspunsului corect în întrebările pe care le-a parcurs deja etc. Din aceste motive, în multe cazuri răspunsurile greşite sunt alese cu frecvenţe diferite unele de altele.

Pornind de la ideile teoriei răspunsului la itemi, Murphy şi Davidshofer (1987) recomandă construirea funcţiei caracteristice a fiecărui răspuns propus. Această funcţie are ca valoare, la fiecare nivel al caracteristicii măsurate de test, probabilitatea ca răspunsul respectiv să fie ales. Dacă itemul este bine construit, atunci funcţia caracteristică a răspunsului corect va fi nedescrescătoare, în timp ce funcţiile caracteristice ale răspunsurilor greşite vor fi necrescătoare; nici una dintre funcţiile caracteristice nu va fi constantă pe întreaga mulţime de definiţie.

Exemplu

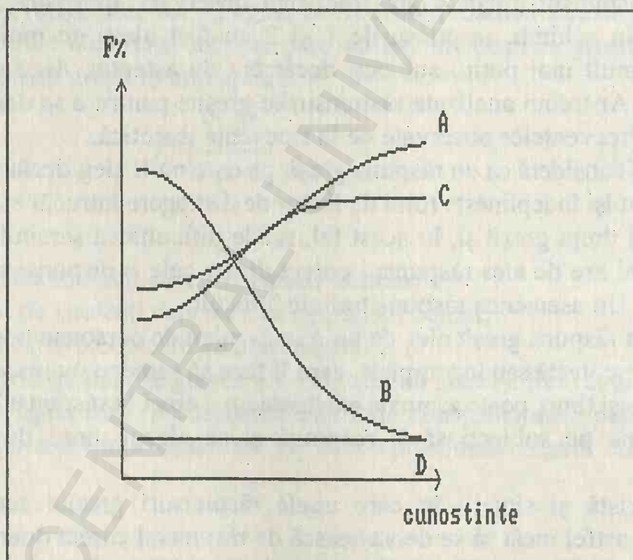


Figura 12.6.1.2.1
Curbele caracteristice ale răspunsurilor la un item cu alegere multiplă

În figura 12.6.1.2.1 sunt prezentate funcţiile caracteristice ale răspunsurilor A (corect), B, C şi D ale unui item cu alegere multiplă dintr-un

test de cunoştinţe. Se observă că numai răspunsul greşit B are funcţia caracteristică descrescătoare. Răspunsul C este ales mai frecvent de subiecţi bine pregătiţi; este posibil ca el să nu fie formulat clar, deci trebuie verificat. Răspunsul D este ales foarte rar şi frecvenţa alegerii sale nu depinde de nivelul cunoştinţelor celor testaţi. El este ineficient ca factor de distragere şi trebuie eliminat.

12.6.1.3. Validitatea itemilor

În general, pentru aprecierea validităţii itemilor se apelează la un criteriu extern, de exemplu la scorurile unui test care măsoară sau prezice acelaşi lucru ca şi testul analizat.

Se spune că un item este valid, dacă el poate fi folosit la evaluarea caracteristicii măsurate de variabila criteriu.

De obicei, pentru estimarea validităţii se calculează coeficientul de corelaţie între scorurile itemului şi valorile variabilei criteriu. Acesta este denumit "indice de validitate" al itemului. Presupunând că variabila criteriu este continuă, tipul coeficientului de corelaţie utilizat va depinde de tipul itemului, mai precis de numărul scorurilor posibile ale itemului.

În cazul itemilor dihotomici, cărora li se acordă un punct dacă sunt rezolvaţi corect şi zero puncte dacă sunt rezolvaţi greşit sau nerezolvaţi, validitatea poate fi măsurată cu ajutorul coeficientului de corelaţie punct-biserială:

$$r_{bis} = \frac{M_C - M_G}{\sigma_y}\sqrt{pq} \qquad (12.6.1.3.1)$$

unde:

p = frecvenţa relativă a celor care au rezolvat corect itemul, egală cu scorul mediu al itemului;

q = 1-p

σ_y = abaterea standard a valorilor variabilei criteriu;

M_C, M_G = media valorilor variabilei criteriu pentru cei care au rezolvat itemul corect, respectiv pentru cei care au greşit sau nu l-au rezolvat.

Corelaţia punct-biserială este un caz particular al corelaţiei produs-moment. Ea corespunde situaţiei când una dintre variabile ia doar două valori.

Dacă testul este administrat cu limită de timp, atunci vor fi luate în considerare pentru calculul corelaţiei numai persoanele care au parcurs itemul.

Verificarea semnificaţiei statistice a coeficientului de corelaţie punct-biserială se face folosind testul t (Student):

$$t = \frac{r_{bis}}{\sqrt{1-r_{bis}^2}}\sqrt{N-2} \qquad (12.6.1.3.2)$$

unde N reprezintă numărul persoanelor care au parcurs itemul.

Itemul este considerat valid, dacă $t > t_{N-2,\infty}$ $t_{N-2,\alpha}$ fiind valoarea citită din tabelele legii t, care corespunde pragului de semnificaţie ales α şi la N-2 grade de libertate.

Se mai poate studia validitatea itemului prin analizarea diferenţei mediilor M_C şi M_G. Presupunând că în cele două grupe determinate de valorile scorului la item (cei cu scorul 1 şi cei cu scorul 0) variabila criteriu este repartizată normal şi are aceeaşi dispersie, se calculează valoarea t folosind formula:

$$t_{dif} = \frac{M_C - M_G}{\sigma_y}\sqrt{Npq} \qquad (12.6.1.3.3)$$

şi se compară cu valoarea $t_{N-2,\alpha}$ citită din tabele. Dacă $t_{dif} > t_{N-2,\alpha}$, atunci se consideră că itemul este valid.

Pentru itemii al căror scor poate lua mai mult de două valori, validitatea poate fi estimată cu ajutorul coeficientului de corelaţie liniară între scorurile la item şi valorile variabilei criteriu.

Interpretarea valorii acestui coeficient se face la fel ca în cazul coeficientului de corelaţie punct-biserială, utilizând testul t.

Atunci când este măsurată printr-un coeficient de corelaţie liniară, validitatea itemului arată cât de puternică este legătura liniară dintre scorurile itemului şi valorile variabilei criteriu.

Uneori este foarte greu sau chiar imposibil de găsit o variabilă criteriu pentru validarea itemilor unui test. În cazul testelor de cunoştinţe asemenea situaţii sunt frecvente.

Dar, dacă specialiştii apreciază că un test, în ansamblu, este valid, adică măsoară cunoştinţele din domeniul pentru care a fost construit, atunci scorurile testului pot fi utilizate, în locul valorilor unei variabile criteriu, la validarea itemilor care compun testul. Este adevărat că rezultatele vor fi distorsionate de faptul că în calculul scorului total al testului se ia în considerare şi scorul itemului, dar, dacă numărul itemilor din test este mare, influenţa fiecărui item

este redusă. Vor fi consideraţi valizi itemii care măsoară acelaşi lucru ca şi întregul test. În acest caz denumirea de indice de validitate este înlocuită ce aceea de **indice de discriminare şi consistenţă internă** (Davis, 1955). Totuşi, *Standards for Educational and Psychological Tests* (1974) consideră că asemenea coeficienţi nu măsoară validitatea itemilor.

12.6.1.4. Indicele de discriminare al unui item

Noţiunea de putere de discriminare a unui item vine să o completeze pe cea de validitate a itemului. Pentru estimarea ambelor este necesară existenţa unei variabile criteriu. Uneori - mai frecvent în cazul testelor de cunoştinţe - drept variabilă criteriu poate servi testul din care face parte întrebarea.

În general, se apreciază că un item are putere de discriminare dacă examinaţii cu valori mari la variabila criteriu au şanse mai mari să rezolve itemul decât cei care au valori scăzute. O asemenea comportare este de aşteptat în cazul itemilor valizi, ale căror scoruri sunt corelate liniar şi pozitiv cu valorile variabilei criteriu. Din acest motiv, în literatura de specialitate problema validităţii itemilor şi cea a puterii lor de discriminare se confundă uneori, deşi metoda utilizată pentru aprecierea puterii de discriminare nu garantează întotdeauna validitatea itemului.

O posibilitate de evaluare a puterii de discriminare a unui item o constituie *metoda grupelor contrastante* (Murphy şi Davidshofer, 1987), aplicabilă atunci când testul a fost administrat unui număr mare de persoane. Din mulţimea celor examinaţi se formează două loturi:

- grupa "bună", compusă din cei care au cele mai mari valori la variabila criteriu;
- grupa "slabă", alcătuită din cei care la variabila criteriu au obţinut valorile cele mai mici.

Dacă valorile variabilei după care se face împărţirea în grupele contrastante sunt repartizate conform legii normale, grupele vor conţine câte 27% din totalul persoanelor. Dacă repartiţia frecvenţelor valorilor se abate mult de la legea normală, se vor forma grupele din câte 33% dintre persoane. În general, procentele persoanelor incluse în grupa bună şi în grupa slabă vor fi aproximativ egale, fiind cuprinse în intervalul [25, 33] (M.J. Allen şi W.M. Yen, după Gregory, 1992).

Pentru fiecare item se calculează indicele de discriminare:

$$D = \frac{C_B}{N_B} - \frac{C_S}{N_S} \qquad (12.6.1.4.1)$$

unde:

N_B, N_S = numărul de persoane în grupa bună, respectiv în grupa slabă. Dacă testul a fost administrat cu limită de timp, atunci N_B şi N_S vor reprezenta numărul de persoane care au parcurs itemul, în fiecare din cele două grupe;

C_B, C_S = numărul persoanelor care au rezolvat itemul corect, în grupa bună, respectiv în grupa slabă.

Deci, indicele de discriminare este diferenţa a două frecvenţe relative. El variază între -1 (când toţi cei din grupa slabă au rezolvat itemul corect şi toţi cei din grupa bună au greşit) şi +1 (când itemul a fost rezolvat corect de toţi cei din grupa bună şi greşit de toţi cei din grupa slabă). Cu cât indicele de discriminare este mai ridicat, cu atât itemul reuşeşte mai bine să diferenţieze între cei "buni" şi cei "slabi".

Valoarea indicelui de discriminare, *D*, este dependentă de cea a indicelui de dificultate, *p*. Atunci când toţi examinaţii rezolvă corect itemul (p=1) sau când nimeni nu răspunde corect (p=0), itemul nu diferenţiază între cei buni şi cei slabi. Dacă valoarea *p* este apropiată de 0 sau de 1 capacitatea de discriminare a itemului este redusă. Cel mai mare potenţial pentru a fi buni discriminatori îl au itemii al căror indice de dificultate este apropiat de 0.5.

Ori de câte ori se obţine o valoare negativă a indicelui de discriminare la un item cu alegere multiplă, este necesar să se revadă răspunsurile propuse. Când unul sau mai multe dintre răspunsurile greşite par foarte plauzibile unui examinat care are un nivel înalt al variabilei măsurate de test (inteligenţă, cunoştinţe etc.) şi când recunoaşterea răspunsului corect depinde de un element extrem de subtil, este posibil ca examinaţii cu un nivel mediu al variabilei evaluate să fie penalizaţi. La un test de cunoştinţe, de exemplu, cei care nu cunosc nimic şi aleg la întâmplare, având şanse să nimerească răspunsul corect, pot să obţină un punctaj mai bun decât cei care au cunoştinţe parţiale şi sunt derutaţi de modul în care sunt formulate răspunsurile propuse.

În cazul întrebărilor dihotomice, atunci când numărul persoanelor examinate este mai mare de 20 şi efectivul celor două grupe este acelaşi ($N_B=N_S$), valoarea indicelui D permite următoarele aprecieri (Guilbert, 1987):

$D \geq 0.35$	item excelent;
$0.25 \leq D < 0.35$	item bun;
$0.15 \leq D < 0.25$	item la limită - de revizuit;
$D < 0.15$	item slab - de reexaminat atent şi, probabil, de eliminat.

Un item poate avea un indice de discriminare convenabil, fără a avea o validitate bună. Un asemenea caz corespunde următoarei situaţii: un item dintr-un test de cunoştinţe pare uşor celor care au un nivel mediu al cunoştinţelor; în realitate, el conţine o greşeală de formulare care face ca două dintre răspunsurile propuse să fie la fel de plauzibile. Examinaţii bine pregătiţi au dubii în legătură cu răspunsul corect şi din acest motiv greşesc mai frecvent decât ceilalţi. Curba de regresie a scorurilor la item faţă de scorurile variabilei criteriu este prezentată în figura 12.6.1.4.1.

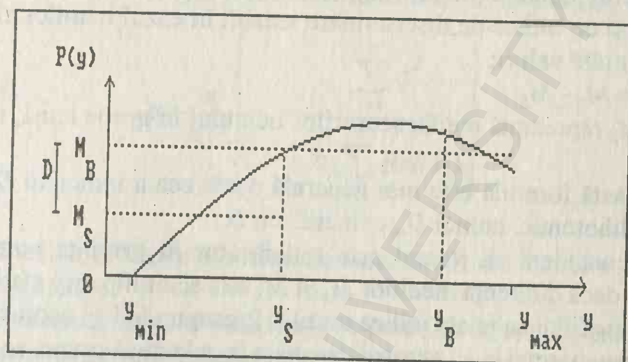


Figura 12.6.1.4.1
Un item care discriminează bine dar nu este valid

Ţinând cont de faptul că indicele de discriminare se exprimă cu ajutorul a două frecvenţe relative, pentru interpretarea corectă a valorii sale se poate aplica un test statistic. Astfel, dacă efectivele celor două grupe sunt mari (N_B>30 şi N_S>30), atunci compararea frecvenţelor relative $p_B=C_B/N_B$ şi $p_S=C_S/N_S$ se poate face cu ajutorul testului z. Se calculează valoarea z cu formula:

$$z = \frac{p_B - p_S}{\sqrt{p(1-p)(\frac{1}{N_B}+\frac{1}{N_S})}} \qquad (12.6.1.4.2)$$

unde:

$$p = \frac{C_B+C_S}{N_B+N_S} \qquad (12.6.1.4.3)$$

şi se compară cu valoarea z_α citită din tabelul legii normale, corespunzătoare pragului de semnificaţie α ales. Dacă $z>z_\alpha$ atunci se apreciază că puterea de discriminare a itemului este satisfăcătoare.

O deficienţă a metodei prezentate pentru calculul indicelui de discriminare constă în faptul că elimină din prelucrare mai mult de o treime din totalul celor examinaţi. În plus, indicele de discriminare astfel calculat nu ne spune cum diferenţiază itemul între persoanele din fiecare grupă, ci doar dacă, în cadrul întregului eşantion, face deosebire între cei mai buni şi cei mai slabi.

Analog indicelui de discriminare D, utilizat pentru itemii dihotomici, se poate defini un indice de discriminare valabil în cazul itemilor al căror scor admite mai multe valori:

$$D_m = M_B - M_S \qquad (12.6.1.4.4)$$

unde M_B şi M_S reprezintă media scorurilor itemului în grupa bună, respectiv în cea slabă.

Această formulă este mai generală decât cea a indicelui D; dar, dacă itemul este dihotomic, atunci D_m coincide cu D.

Se consideră că itemul este satisfăcător în privinţa puterii sale de discriminare dacă diferenţa mediilor M_B şi M_S este semnificativă statistic. Pentru compararea mediilor se poate utiliza testul t. Presupunând că scorurile itemului sunt repartizate normal şi au aceeaşi dispersie în cele două grupe, se calculează:

$$t = \frac{M_B - M_S}{\sqrt{\dfrac{(N_B-1)\sigma_B^2+(N_S-1)\sigma_S^2}{N_B+N_S-2}}}\sqrt{\frac{N_B+N_S}{N_B N_S}} \qquad (12.6.1.4.5)$$

unde σ_B^2 şi σ_S^2 reprezintă dispersia scorurilor itemului în grupa bună, respectiv în grupa slabă.

Dacă t este mai mare decât valoarea $t_{m,\alpha}$ citită din tabelele legii t pentru $m=N_B+N_S-2$ grade de libertate şi pragul de semnificaţie α, se consideră că itemul are o putere de discriminare satisfăcătoare.

Indicele de discriminare D_m, la fel ca şi D, arată doar în ce măsură o întrebare diferenţiază între o grupă bună şi una slabă. El nu spune nimic despre felul în care itemul discriminează între indivizi care au niveluri diferite ale variabilei măsurate sau prezise de test.

12.6.1.5. Indicele de fidelitate al itemului

Acest indice serveşte la aprecierea contribuţiei pe care şi-o aduce itemul la asigurarea unei bune consistenţe interne a testului.

Conform formulei (8.4.2.2), indicele de fidelitate este dat de produsul dintre abaterea standard a scorurilor sale, σ_i, şi coeficientul de corelaţie liniară dintre scorurile itemului şi scorurile testului, $\rho(Y_i,X)$:

$Rel(Y_i)=\sigma_i\rho(Y_i,X)$

Din formula de calcul a coeficientului α al lui Cronbach, (8.4.2.3):

$$\alpha=\frac{n}{n-1}\left(1-\frac{\sum_{i=1}^{n}\sigma_i^2}{\sigma_X\sum_{i=1}^{n}Rel(Y_i)}\right)$$

se observă că testele alcătuite din itemi cu valori mari ale indicilor de fidelitate au o consistenţă mare.

Este dificil de apreciat contribuţia pe care şi-o aduce dispersia scorurilor itemului la mărimea coeficientului de consistenţă internă, întrucât în expresia coeficientului α abaterea standard a scorurilor itemului, σ_i, apare atât la numărătorul, cât şi la numitorul raportului care se scade din 1. Dar, pentru obţinerea unei valori cât mai mari a coeficientului α este important ca itemii care vor fi introduşi în test să aibă corelaţii liniare mari, pozitive, cu scorul total al testului.

12.6.1.6. Observaţii asupra analizei de itemi

Analiza itemilor unui test se poate face conform următorului algoritm:

(1) Se calculează indicele de dificultate pentru toţi itemii şi se elimină itemii rezolvaţi de toate persoanele şi cei nerezolvaţi de nici un subiect.

(2) Se încearcă depistarea cauzelor pentru care unii indici de dificultate sunt foarte mari sau foarte mici şi, pe această bază, se elimină itemii cu greşeli.

(3) Pentru itemii cu răspunsuri la alegere, se analizează răspunsurile incorecte şi se elimină itemii în care unele răspunsuri greşite propuse au fost alese de foarte mulţi sau de foarte puţini subiecţi.

(4) Se aplică un procedeu al analizei de itemi determinat de condiţiile pe care trebuie să le îndeplinească testul care se construieşte:

- Dacă testul trebuie să discrimineze între două grupe contrastante, mai precis, să furnizeze scoruri care să difere mult de la o grupă la alta, atunci se calculează pentru fiecare item indicele de discriminare şi se elimină itemii necorespunzători din acest punct de vedere.
- Dacă testul trebuie să se comporte într-un anumit fel faţă de anumit criteriu (să coreleze liniar pozitiv sau negativ ori, din contră, să nu coreleze semnificativ), atunci se vor elimina itemii care au o altfel de comportare.

(5) Se elimină itemii cu indicele de dificultate necorespunzător scopului pe care trebuie să îl îndeplinească testul: în general, se elimină itemii foarte uşori şi cei foarte grei, dar, dacă testul trebuie să identifice persoanele cu niveluri foarte scăzute sau foarte ridicate ale unei caracteristici, se vor reţine doar itemii uşori, respectiv dificili.

(6) Se calculează coeficientul de corelaţie între scorurile testului şi scorurile itemului şi se elimină itemii cu corelaţii nesemnificative sau negative.

Din itemii rămaşi vor fi selectaţi cei care vor alcătui testul. Pentru testul rezultat se vor studia atât caracteristicile psihometrice (fidelitate, validitate, dificultate, putere de discriminare), cât şi lungimea testului, tipurile itemilor componenţi şi, în funcţie de acestea, timpul de rezolvare a testului şi costurile administrării sale.

12.6.2. Metode de selectare a itemilor pentru construirea unui test psihologic

Metoda intuitivă (raţională) se utilizează pentru a obţine un test care să măsoare cu precizie diferenţele individuale în privinţa unei trăsături comune sau a unui atribut care prezintă interes ştiinţific şi care nu este măsurat adecvat prin nici o tehnică existentă (Edwards, 1970). Se deduce că această metodă urmăreşte formarea unui test pentru care sunt importante validitatea relativă la construct şi consistenţa internă.

Se porneşte de la o mulţime de itemi care descriu comportamente specifice constructului la care se referă testul, formulaţi pe baza intuiţiei autorului testului. De aici provine denumirea de "metodă intuitivă". În unele lucrări, denumirea de "metodă raţională" se foloseşte ca un sinonim pentru

"metoda intuitivă", în alte lucrări este considerată "metodă raţională" doar cea care utilizează procedee statistice pentru selectarea itemilor din colecţia existentă (Henry, Sanford, 1972)

Pentru a putea aplica formule statistice, itemii sunt administraţi unui lot format din 200-400 de persoane (Edwards, 1970) şi se calculează coeficientul de consistenţă internă. Dacă acesta are o valoare prea scăzută se încearcă să se formeze testul dintr-un număr mai restrâns de itemi.

În general se utilizează două căi de selecţie a itemilor, ambele bazate pe scorurile acestora în cadrul eşantionului:

- metoda consistenţei interne;
- metoda teoretică.

A. Metoda consistenţei interne are mai multe variante. Vom prezenta câteva dintre acestea:

- Dacă numărul de itemi este mai mare decât 60, se elimină itemii care au covarianţe scăzute cu scorul total al testului, iar dacă numărul de itemi este mai mic decât 20, se adaugă itemi noi (Edwards, 1970).
- Sunt eliminaţi itemii care au coeficienţi de corelaţie scăzuţi cu scorul total al testului (Henry, Sanford, 1972).
- Sunt îndepărtaţi itemii care au o putere de discriminare scăzută, grupele contrastante utilizate pentru calculul indicelui de discriminare fiind formate pe baza scorurilor testului (Anastasi, 1976).

Deficienţa acestor metode constă în faptul că nu asigură selectarea acelor itemi care ar furniza cea mai mare valoare posibilă a coeficientului de fidelitate. În metodele consistenţei interne fiecare item este comparat cu întregul test şi este posibil ca un item necorespunzător să provoace eliminarea unor itemi buni.

B. Metoda teoretică (Shackleton, Fletcher, 1984) grupează itemii şi reţine itemii unei clase omogene.

Pentru grupare se foloseşte un algoritm al analizei de clusteri.

Prin acest algoritm la fiecare pas se construieşte pe mulţimea de itemi o partiţie. În cadrul fiecărei partiţii, clasa care se referă la constructul măsurat de test trebuie identificată prin analiza conţinutului itemilor. Se formează testul din itemii clasei respective şi se calculează coeficientul de fidelitate al testului. Deci, la fiecare pas al algoritmului există câte un test. Dintre acestea se alege cel care are o fidelitate bună şi este compus dintr-un număr suficient de mare de itemi ca să poată diferenţia între subiecţi.

În privinţa modului de validare a testului construit prin metoda raţională, părerile sunt contradictorii.

Edwards (1970) atrage atenţia că un asemenea test nu trebuie corelat cu nici un criteriu, deoarece un test se construieşte prin metoda raţională doar atunci când nu există nici o altă posibilitate de a măsura trăsătura la care se referă acesta.

Anastasi (1976) propune să se studieze corelaţia dintre testul construit şi alte teste similare existente, pentru a vedea dacă toate măsoară acelaşi lucru. Valorile coeficienţilor de corelaţie trebuie să fie mari, dar nu foarte mari, căci dacă se identifică o corelaţie puternică, fără ca noul test să aibă avantajul de a fi mai uşor de administrat, se poate considera că el nu este decât o dublură a testului existent şi nu prezintă nici un interes.

D.T. Campbell (Anastasi, 1976) apreciază că pentru a verifica validitatea relativă la construct a testului trebuie să se studieze corelaţiile acestuia cu teste care se referă la variabile psihice despre care se consideră că au legătură cu constructul măsurat de test (validitatea convergentă) şi cu teste care vizează variabile psihice despre care se presupune că nu au legătură cu constructul măsurat de test (validitatea discriminantă).

Metoda analizei criteriului intern (metoda analizei factoriale) (Mischel, 1968; Henry, Sanford, 1972; Kline, 1993) serveşte la construirea unui test destinat să măsoare un anumit factor, care se consideră că reprezintă o variabilă de personalitate fundamentală.

Sunt scrişi mai mulţi itemi. Aceştia sunt administraţi unui lot de subiecţi împreună cu alte măsuri (teste, itemi etc.) care au fost identificate în cadrul unor cercetări anterioare ca având legătură cu variabila respectivă. Matricea intercorelaţiilor scorurilor obţinute pentru toate măsurătorile efectuate este prelucrată prin analiză factorială. Se identifică factorul pe care trebuie să îl măsoare testul şi se reţin numai itemii care au saturaţii mari în factorul respectiv. Dacă este necesar, se schimbă modul de cotare a itemilor astfel încât toate saturaţiile să fie pozitive.

Trebuie menţionat că factorul comun unei grupări de itemi nu corespunde, în mod obligatoriu, unei caracteristici "reale" sau "primare" a persoanelor examinate (Mischel, 1968). Factorii identificaţi prin analiză factorială sunt dependenţi de itemii prelucraţi statistic, de subiecţii care au răspuns la test şi de metodele aplicate în cadrul analizei factoriale. Variabilele puse în evidenţă de analiza factorială reflectă doar corelaţiile dintre mai multe instrumente de măsură particulare (itemii testului).

Prin metoda analizei factoriale este mai uşor să se construiască deodată mai multe teste decât unul singur. Aceasta deoarece pasul iniţial al analizei factoriale are tendinţa de a evidenţia un factor general şi mai mulţi factori bipolari, iar prin rotaţia factorilor se reduce varianţa factorului general (Kline,

1993). Prin urmare, de la început se vor selecta itemi care să măsoare mai mulţi factori. Aceştia vor fi administraţi unui lot format cam din 200 de persoane sau, în orice caz, care nu este mai mic de 100 de indivizi. Kline (1993) recomandă ca numărul itemilor prelucraţi să fie de două ori mai mare decât cel al itemilor care vor fi păstraţi în teste, iar numărul subiecţilor să fie de 3 ori mai mare decât numărul itemilor introduşi în analiza factorială. Astfel, de exemplu, dacă se vor construi 5 scale a câte 20 de itemi, atunci vor trebui administraţi 2*5*20=200 de itemi, unui lot format din 3*200=600 de persoane.

În ceea ce priveşte tipul de rotaţie utilizat în analiza factorială, se va aplica o rotaţie ortogonală dacă există argumente teoretice în favoarea ideii că variabilele măsurate de testele care vor fi construite sunt necorelate liniar între ele; în caz contrar, se va aplica o rotaţie oblică. Pentru fiecare test vor fi selectaţi numai itemii care au saturaţii mari într-un singur factor.

Prin această metodă se obţin teste care măsoară câte un singur factor. Pentru a afla ce reprezintă acest factor, se va studia modul în care corelează scorurile testului cu diverse teste despre care se ştie ce măsoară.

Această metodă se foloseşte atunci când se doreşte să se obţină teste cu o validitate relativă la construct bună.

Metoda exploratorie (Ozer, Reise, 1994) se foloseşte pentru construirea testelor destinate să evalueze constructe care la început sunt incomplet înţelese. În paralel cu construirea testului se dezvoltă şi teoria care stă la baza sa. De fapt, această metodă transpune în practică spirala inductiv-ipotetico-deductivă a lui Cattell.

Se începe cu o idee nefinisată despre un construct de personalitate şi se scriu cât mai mulţi itemi referitori la el. Aceştia sunt administraţi unui lot de subiecţi. Răspunsurile obţinute nu sunt folosite pentru a studia şi îmbunătăţi calităţile psihometrice ale testului, ci pentru a genera teorii noi. Sunt scrişi alţi itemi, potriviţi noii idei despre construct, iar răspunsurile obţinute servesc la o nouă evaluare a teoriei. Procesul se repetă până când se ajunge la un nivel satisfăcător de convergenţă a răspunsurilor. Itemii respectivi vor constitui testul.

Metoda criteriului extern se utilizează de obicei în cazul testelor care trebuie să aibă o validitate predictivă ridicată.

Este foarte important ca itemii introduşi în prelucrare să se refere la aspecte cât mai diverse, să formeze o colecţie cât mai eterogenă. Alegerea itemilor care vor intra în componenţa testului se bazează pe relaţiile, determinate empiric, dintre itemii testului şi o măsură particulară folosită drept criteriu. Din acest motiv, metoda este denumită şi "empirică" (Edwards, 1970; Henry, Sanford, 1972).

De obicei se studiază validitatea fiecărui item relativ la acel criteriu şi se elimină itemii care nu sunt valizi.

Uneori, criteriul împarte indivizii în două clase contrastante: unii, pentru care o anumită trăsătură, caracteristică sau aptitudine este foarte accentuată, formează grupul "criteriu", iar ceilalţi constituie grupul "de control". În acest caz se studiază puterea de discriminare a itemilor şi se elimină itemii care nu discriminează între grupurile contrastante.

Cele două grupuri contrastante trebuie să conţină, împreună, cel puţin 200 de persoane, dacă toate sunt de acelaşi sex, sau cel puţin 200 de bărbaţi şi 200 de femei. În cel de-al doilea caz, se selectează doar itemii care discriminează între grupurile contrastante pentru ambele sexe.

Adesea, criteriul care stă la baza împărţirii persoanelor în cele două grupuri este multidimensional. Din acest motiv, este posibil ca testul format din itemii reţinuţi să nu diferenţieze suficient de bine între grupuri. Sau este posibil ca fiecare item să se refere la un alt aspect al criteriului; aceasta face ca intercorelaţiile itemilor să fie reduse şi, deci, consistenţa internă a testului să fie scăzută. Pentru a înţelege ce măsoară testul (care sunt factorii care discriminează între grupurile contrastante) se pot prelucra prin analiză factorială itemii reţinuţi împreună cu itemi ai altor teste, care se ştie ce măsoară, şi apoi se vor identifica factorii obţinuţi.

Metoda analizei constructului (Oosterveld, Vorst, 1995) se bazează pe corelaţiile itemilor cu diverse variabile care, conform teoriei ce stă la baza elaborării testului, au legături sau, din contră, nu trebuie să aibă legături cu acest construct. Itemii ale căror corelaţii sunt necorespunzătoare se elimină.

Această metodă se aseamănă cu metoda empirică, doar că aici se folosesc mai multe criterii, inclusiv unele cu care itemii trebuie să nu coreleze semnificativ. Aplicarea metodei constructului necesită existenţa unei reţele nomologice care să includă constructul măsurat de test.

Alte metode de construire a testelor

Algoritmii pe care îi prezentăm în continuare construiesc un test pornind de la o colecţie de itemi dihotomici şi de la un item fixat, apreciat ca fiind reprezentativ pentru ceea ce testul trebuie să măsoare sau să prezică.

La fiecare pas se adaugă un item testului existent în acel moment. Itemului adăugat (care se alege astfel încât să maximizeze coeficientul de consistenţă internă, respectiv un coeficient de discriminare între grupele contrastante) i se impune să aibă funcţia de regresie faţă de scorurile testului existent nedescrescătoare. În plus, se ţine cont şi de următoarea observaţie:

Este posibil ca fiecare dintre itemii $i_1, i_2,..., i_n$ să aibă funcţia de regresie monotonă faţă de testul format din ei, itemul i_{n+1} să aibă funcţia de regresie monotonă faţă de testul alcătuit din itemii $i_1, i_2,..., i_n, i_{n+1}$, dar să existe cel puţin un item $i_j, j \in \{1, 2, ..., n\}$ a cărui funcţie de regresie faţă de testul format din itemii $i_1, i_2,..., i_{n+1}$ să nu fie monotonă.

Din acest motiv nu se adaugă un item testului decât după ce se verifică dacă toţi itemii existenţi în test au funcţii de regresie monotone faţă de testul care s-ar forma prin adăugarea itemului.

Algoritmii aplică următoarele formule de calcul:

- Scorul testului se obţine prin însumarea cotelor itemilor componenţi.
- Pentru determinarea funcţiei de regresie a unui item faţă de test se aplică metoda celor mai mici pătrate folosind polinoame de gradul al III-lea. Concret, dacă testul este alcătuit din n itemi: $i_1, i_2,..., i_n$ scorurile posibile fiind $0, 1, ..., n$ şi dacă pentru fiecare item i_k $(k=1,2,...,n)$ notăm cu p_k funcţia:

$p_k : \{0, 1,..., n\} \to [0,1]$

$$p_k(t) = \begin{cases} \text{frecvenţa relativă a persoanelor care obţin un punct la itemul } i_k \text{ din totalul celor care au realizat } t \text{ puncte la test, } t=1, 2,..., n \\ 0, t=0 \end{cases}$$

atunci se determină funcţia de regresie a itemului i_k faţă de test, c_k, care îndeplineşte următoarele condiţii:

$c_k : [0,n] \to [0,1]$

$c_k \in \{a_0+a_1x+a_2x^2+a_3x^3 \mid a_0,a_1,a_2,a_3 \in \mathbb{R}\}$

$$\sum_{i=0}^{n} [p_k(i) - c_k(i)]^2$$

ia valoarea minimă.

Se observă că prin condiţiile impuse nu s-au eliminat dreptele de regresie. Ele corespund cazului când c_k este un polinom de gradul întâi pe intervalul $[0,n]$. Dar, pentru unii itemi i_k, se poate găsi o funcţie c_k care să aproximeze punctele $p_k(0), p_k(1),..., p_k(n)$ mai bine decât dreapta de regresie şi, în plus, să imite mai bine alura funcţiei p_k, punând în evidenţă eventualele schimbări de monotonie, ceea ce polinoamele de gradul întâi nu sunt capabile să facă.

A. Un algoritm de construire a unui test cu o consistenţă internă ridicată (Albu, Pitariu, 1992b)

Algoritmul pe care îl prezentăm în continuare construieşte un test psihologic prin extragerea itemilor dintr-o mulţime de itemi dihotomici M, astfel încât:

• funcţia de regresie a fiecărui item al testului este nedescrescătoare faţă de test;

• la testul existent la un moment dat se adaugă acel item care produce cel mai mare coeficient de fidelitate Kuder-Richardson, ρ_{20}.

Algoritmul se compune din următorii paşi:

Pasul 1 : M := $\{i_1, i_2,..., i_n\}$ este mulţimea de itemi;
i_0 este itemul de pornire;
$s := \{i_0\}$ este testul construit
$k := 1$ este numărul de itemi ai testului s

Pasul 2 : $P = \emptyset$

Pentru fiecare item $i_j \in$M se determină funcţia de regresie a itemului i_j faţă de testul $s \cup \{i_j\}$ şi, dacă aceasta este nedescrescătoare, se pune $P := P \cup \{i_j\}$.

Pasul 3 : Dacă $P=\emptyset$, se trece la pasul 8.

Pasul 4 : Se pune $KRMAX := 0$ şi $JMAX := 0$

Pasul 5 : Pentru fiecare $i_j \in P$:

• se calculează, folosind scorurile testului $s_j := s \cup \{i_j\}$ coeficientul de fidelitate ρ_{20} al testului s_j: $KR(s_j)$;

• dacă $KR(s_j) > KRMAX$, se pune $KRMAX := KR(s_j)$) şi $JMAX := j$

Pasul 6 : Pentru fiecare item $i_m \in s \cup \{i_{JMAX}\}$:

• se determină funcţia de regresie faţă de testul $s \cup \{i_{JMAX}\}$;

• dacă aceasta nu este nedescrescătoare, se elimină itemul i_{JMAX} din mulţimea P :

$$P := P \setminus \{i_{JMAX}\}$$

şi se trece la pasul 3.

Pasul 7 : Se adaugă itemul i_{JMAX} testului s şi se elimină din M :

$$s := s \cup \{i_{JMAX}\}$$
$$\mathrm{M} := \mathrm{M} \setminus \{i_{JMAX}\}$$
$$k := k+1$$

Dacă M $\neq \emptyset$ se trece la pasul 2.

Pasul 8 : STOP

S-a obţinut testul s cu k itemi.

Algoritmul poate fi modificat, introducând condiţia de oprire dacă valoarea coeficientului de fidelitate se află sub un anumit prag sau atunci când valorile coeficienţilor ρ_{20} scad de la un pas la altul.

B. Un algoritm de construire a unui test care să discrimineze bine între două grupe contrastante (Albu, Pitariu, 1991b)

Acest algoritm serveşte la construirea unui test care să discrimineze cât mai bine între două grupe contrastante, G_1 şi G_2, apreciate a fi foarte deosebite în privinţa variabilei la care se referă testul.

El urmăreşte ca:

• itemii introduşi în test să aibă funcţia de regresie faţă de test nedescrescătoare;

• scorurile medii ale testului în cele două grupe, G_1 şi G_2, să fie cât mai diferite posibil;

• în cadrul fiecărei grupe, scorurile testului să fie cât mai apropiate de scorul mediu al grupei.

Ultimele două condiţii de mai sus sus au fost formulate matematic cu ajutorul funcţiei **C**:

C : S -> R

$$C(s)=\frac{m_1-m_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \qquad (12.6.2.1)$$

unde :

S = mulţimea tuturor testelor care se pot forma cu itemii mulţimii **M**;

s = un test din **S**;

pentru i=1,2 :

n_i = volumul grupei G_i;

m_i = scorul mediu al testului s în grupa G_i;

σ_i^2 = dispersia scorurilor testului s în grupa G_i.

Înainte de aplicarea algoritmului se verifică modul de cotare a itemilor, astfel încât fiecare item să aibă media scorurilor mai mare în grupa G_1.

Algoritmul se compune din următorii paşi:

Pasul 1 : M := $\{i_1, i_2,..., i_n\}$ este mulţimea de itemi;

s := Ø este testul construit

k := 0 este numărul de itemi ai testului s

Pasul 2 : Se pune P:=Ø

Pentru fiecare item $i_j \in$ **M**:

• dacă s = Ø se pune $P := P \cup \{i_j\}$;

• dacă $s \neq$ Ø, se determină funcţia de regresie a itemului i_j faţă de testul $s \cup \{i_j\}$ şi, dacă aceasta este nedescrescătoare, se pune $P := P \cup \{i_j\}$.

Pasul 3 : Dacă $P=\emptyset$, se trece la pasul 7.
Pasul 4 : Se pune $CMAX := 0$ şi $JMAX := 0$

Pentru fiecare item $i_j \in P$:

- se calculează, folosind scorurile testului $s_j := s \cup \{i_j\}$:

$$C(s_j) = \frac{m_1 - m_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- dacă $C(s_j) > CMAX$, se pune $CMAX := C(s_j)$ şi $JMAX := j$

Pasul 5 : Pentru fiecare item $i_m \in s \cup \{i_{JMAX}\}$:

- se determină funcţia de regresie faţă de testul $s \cup \{i_{JMAX}\}$;
- dacă aceasta nu este nedescrescătoare, se elimină itemul i_{JMAX} din mulţimea P :

$$P := P \setminus \{i_{JMAX}\}$$

şi se trece la pasul 3.
Pasul 6 : Se adaugă itemul i_{JMAX} testului s şi se elimină din **M** :

$$s := s \cup \{i_{JMAX}\}$$
$$\mathbf{M} := \mathbf{M} \setminus \{i_{JMAX}\}$$
$$k := k+1$$

Dacă $\mathbf{M} \neq \emptyset$ se trece la pasul 2.
Pasul 7 : STOP

S-a obţinut testul s cu k itemi.

12.7. EXPERIMENTAREA SUPLIMENTARĂ A TESTULUI

Efectuarea unei verificări suplimentare a testului, folosind un eşantion de persoane asemănător celui folosit pentru construirea sa, este necesară ori de câte ori alcătuirea testului s-a realizat prin selecţia itemilor dintr-o bancă de itemi pe baza răspunsurilor date de subiecţi (deci itemii nu au fost aleşi de experţi). Adică, s-a format un test care este "cel mai bun posibil" sau unul dintre "cele mai bune posibil" pentru eşantionul de persoane folosit, ţinând cont de itemii care au fost administraţi şi de criteriul de optim utilizat (de exemplu, s-a urmărit maximizarea fidelităţii sau a validităţii relative la criteriu). La formarea testului şi-au adus contribuţia şi diverşi factori legaţi de şansa care a acţionat în eşantionul de persoane care a participat la construirea testului dar care s-ar putea

să nu mai fie prezenţi în alte ocazii de testare, cu alţi subiecţi. Din acest motiv trebuie verificată comportarea testului şi într-un alt lot de persoane.

În literatura de specialitate se recomandă ca după alcătuirea testului să se efectueze o **validare suplimentară** (cross-validation). Prin aceasta se verifică, folosind ecuaţia de regresie a scorurilor testului faţă de un criteriu, dacă testul prezice criteriul într-un eşantion nou de subiecţi la fel de bine ca în eşantionul original (Gregory, 1992). În practică s-a constatat o scădere a valorii coeficientului de validitate prin schimbarea lotului de subiecţi (validity-shrinkage).

Alături de studiul validităţii testului, este bine că în noul eşantion să se cerceteze şi fidelitatea, dificultatea şi sensibilitatea testului.

În această fază de experimentare suplimentară a testului se verifică nu numai calităţile psihometrice ale acestuia ci şi felul în care au fost redactate instrucţiunile de administrare şi au fost editaţi itemii.

Pentru a se observa eventualele deficienţe, este util să li se dea subiecţilor un chestionar de opinie, pe care să îl completeze imediat după ce au terminat de răspuns la test. Un asemenea chestionar este anonim şi conţine întrebări la care se răspunde prin "da" sau "nu" şi, eventual, se cere justificarea unuia dintre răspunsuri. Exemple de asemenea întrebări sunt următoarele:

- Condiţiile de testare (temperatură, lumină, ventilaţie etc.) au fost satisfăcătoare? (DA/NU)
  Dacă răspunsul este NU, explicaţi ce v-a nemulţumit.
- Instrucţiunile de lucru au fost complete? (DA/NU)
  Dacă răspunsul este NU, ce ar fi trebuit spus în plus?
- Instrucţiunile de lucru au fost clare? (DA/NU)
  Dacă răspunsul este NU, ce a fost neclar?
- Textul itemilor a fost uşor de înţeles? (DA/NU)
  Dacă răspunsul este NU, unde aţi avut dificultăţi?
- Modul în care a trebuit să completaţi răspunsurile a fost simplu? (DA/NU)
  Dacă răspunsul este NU, ce a fost dificil?
- Timpul de lucru a fost suficient? (DA/NU)
- Aţi răspuns la toţi itemii? (DA/NU)
  Dacă răspunsul este NU, de ce?
- Vi s-a părut bun testul? (DA/NU)
  Dacă răspunsul este NU, ce nu v-a plăcut?)

12.8. TRADUCEREA ŞI ADAPTAREA UNUI TEST

Pentru aplicarea unui instrument de măsură într-un grup cultural nou nu este totdeauna suficientă traducerea textului itemilor şi a instrucţiunilor de administrare. Uneori este necesară adaptarea testului, prin înlocuirea unor itemi care nu se potrivesc noii populaţii, alteori, testul trebuie reconstruit. Aceasta, deoarece între populaţia căreia i-a fost destinat testul iniţial şi noua populaţie pot exista deosebiri în ceea ce priveşte (Van de Vijver, Hambleton, 1996):

a. comportamentele asociate constructului măsurat de test;

De exemplu, pentru constructul "pietate filială", comportamentele asociate cu ceea ce se înţelege prin "a fi un fiu bun" (a avea grijă de părinţi, a le satisface cerinţele, a-i trata bine) sunt mult mai numeroase în China decât în Europa de Vest.

Se impune cercetarea modului în care este înţeles constructul în viaţa zilnică şi a comportamentelor asociate lui. Se pot identifica astfel itemii nepotriviţi noii populaţii.

b. obişnuinţa persoanelor cu situaţia de testare ori cu forma particulară a testului;

Lipsa obişnuinţei de a răspunde la teste (în general sau la teste de acest tip) influenţează asupra răspunsurilor la un număr mare de itemi sau chiar la toţi itemii, producând diferenţe ale scorului total al testului între cele două populaţii.

c. răspunsurile date la unii itemi din test, care nu sunt potriviţi noii populaţii.

Această situaţie poate fi identificată observând, pentru persoane din populaţii diferite care au acelaşi scor total la test, dacă există itemi la care răspunsurile sunt diferite. Studiul statistic se bazează în acest caz pe analiza de varianţă, variabila dependentă fiind scorul itemului, iar variabilele independente, populaţia şi scorul total al testului.

După efectuarea unei cercetări comparative, pentru identificarea itemilor la care răspunsurile se deosebesc între populaţii şi depistarea cauzelor, se va decide cum trebuie procedat pentru a putea utiliza testul şi în noua populaţie.

În anul 1993 a fost format un comitet internaţional (International Test Comission), compus din specialişti din diverse domenii ale psihologiei, care, în 1996, au redactat un set de reguli pentru traducerea testelor (Van de Vijver, Hambleton, 1996).

CAPITOLUL 13
INTERPRETAREA SCORURILOR TESTELOR PSIHOLOGICE

13.1. PRINCIPII GENERALE

Scorul obţinut de o persoană la un test, indiferent de formula utilizată pentru calculul acestuia (însumarea scorurilor itemilor, ponderate sau nu, aplicarea formulei de corecţie pentru şansa succesului etc.), se numeşte **cotă brută**.

Valoarea sa nu aduce nici o informaţie despre priceperile sau cunoştinţele subiectului. Faptul că un elev a obţinut 60 de puncte la un test de atenţie concentrată nu ne spune dacă a realizat o performanţă bună, mediocră sau slabă. O cotă brută dobândeşte semnificaţie doar dacă ea este raportată:

- la baremuri fixate

sau

- la scorurile realizate de alţi subiecţi

sau

- la cota brută obţinută de aceeaşi persoană la o testare anterioară.

Prima situaţie este întâlnită mai ales la testele centrate pe criteriu şi la cele folosite în selecţie. Uneori, baremul la care se face raportarea rezultatelor unui test este reprezentat de o singură valoare de scor. Aceasta poate fi un scor de secţionare (cutoff score) sau un scor critic (critical score) (Truxillo, Donahue, Sulzer, 1996).

Scorul de secţionare este un prag care serveşte la împărţirea subiecţilor în două categorii: "acceptaţi" (cei care au un scor superior scorului de secţionare) şi "respinşi" (celelalte persoane). Uneori, scorul de secţionare este stabilit de specialişti, care apreciază nivelul minim de cunoştinţe, deprinderi ori priceperi necesare pentru a obţine succes într-o profesie, la un curs de calificare sau specializare etc. De exemplu, la un test de cunoştinţe, ei decid numărul minim de întrebări la care o persoană trebuie să răspundă corect pentru a fi declarată "admisă". Alteori, scorul de secţionare se fixează în urma administrării testului la două grupe contrastante de persoane, astfel încât majoritatea scorurilor obţinute de grupa "bună" să se situeze într-o parte a sa, iar majoritatea scorurilor din grupa "slabă"în cealaltă parte. Se recurge la fixarea unui scor de secţionare nu numai în cazul selecţiei profesionale, pentru a elimina o parte din candidaţi, ci şi la alte examinări psihologice, pentru a le oferi un feedback subiecţilor, care, de multe ori, sunt mai puţin interesaţi de locul pe care îl ocupă

în clasamentul întocmit pe baza rezultatelor testării, decât de faptul că "au trecut" sau "au căzut" la test (Truxillo, Donahue, Sulzer, 1996).

Scorul critic serveşte la diferenţierea persoanelor care au succes de cele care au insucces într-un anumit domeniu. De obicei stabilirea acestui scor utilizează ecuaţia de regresie liniară dintre criteriu şi predictor, el având rolul de a-i identifica pe subiecţii care au valori mari la criteriu, pe baza scorurilor obţinute la test.Celelalte două modalităţi de interpretare a scorului se întâlnesc la majoritatea testelor normative.

Dar, indiferent de felul în care se procedează pentru a desprinde concluzii din rezultatele unei testări psihologice, este bine să se ţină seama de următoarele recomandări (Gronlund, 1967):

A. Înainte de a interpreta scorurile obţinute la un test se va studia manualul testului, pentru a cunoaşte exact semnificaţia variabilei măsurate sau prezise de test. Simpla denumire a testului nu aduce suficiente informaţii pentru o interpretare corectă. Un "test de aritmetică", de exemplu, poate conţine exerciţii simple de calcul sau probleme de raţionament.

B. Nu se pot face deducţii corecte numai pe baza unui scor de test. Scorul reflectă, pe lângă mărimea variabilei măsurate de test, şi aptitudinile persoanei, influenţele exercitate asupra ei de mediul cultural din care face parte şi de programul instructiv-educativ urmat, adaptarea emoţională, starea de sănătate etc. Un scor scăzut poate fi cauzat, de pildă, de neînţelegerea termenilor utilizaţi de test sau de lipsa motivaţiei.

C. Când se interpretează rezultatele unei testări psihologice se va avea în vedere la ce vor servi concluziile la care se va ajunge. De exemplu, un scor la un test de inteligenţă are semnificaţie diferită, după cum se doreşte să se selecteze elevi pentru a forma o clasă de copii capabili de performanţe superioare ori se urmăreşte să se decidă dacă subiectul trebuie îndrumat spre liceu sau spre o şcoală profesională.

D. Este preferabil ca interpretarea rezultatelor la un test să se bazeze nu pe scorurile observate, ci pe intervalele de încredere ale scorurilor reale. Această recomandare figurează şi în *Standards for Educational and Psychological Tests* din 1974.

E. Orice scor obţinut la un test este afectat, mai mult sau mai puţin, şi de erori de măsurare a căror mărime nu poate fi determinată. Din acest motiv, se va evita să se ia decizii importante pe baza unui singur scor de test.

13.2. INTERPRETAREA SCORURILOR LA TESTELE DE CUNOŞTINŢE INFORMATIVE

Să considerăm, de exemplu, un test de cunoştinţe realizat şi administrat de un profesor pentru o clasă, compus din 50 de itemi care conţin câte un exerciţiu de calcul aritmetic. Pentru fiecare exerciţiu rezolvat corect se acordă un punct, iar cota brută la test a unei persoane este dată de suma cotelor itemilor, deci de numărul de exerciţii rezolvate corect.

Pentru a interpreta rezultatul unui elev, profesorul poate proceda în două moduri:

- să calculeze procentul de itemi rezolvaţi corect de elev din totalul itemilor testului;
- să compare cota brută a elevului cu scorurile celorlalte persoane, pentru a vedea dacă elevul respectiv a obţinut un rezultat dintre cele mai bune, unul mediocru sau unul dintre cele mai slabe.

Prima metodă este utilă doar în cazul testelor de cunoştinţe cu criterii absolute, când procentul de rezolvări corecte arată cât de aproape se află fiecare elev de performanţa minimă necesară. De exemplu, dacă se consideră că pentru promovarea clasei întâi elevii trebuie să fie în stare să rezolve corect cel puţin 45 de exerciţii din test, atunci performanţa minimă necesară este de 45/50*100%=90%, iar un elev care a rezolvat corect doar 20 de exerciţii, adică 20/50*100%=40%, se află cu mult sub nivelul necesar promovării.

Trebuie menţionat, însă, că fixarea şi utilizarea criteriilor absolute pentru a se lua decizii importante asupra celor examinaţi (de exemplu, promovarea clasei) impune o verificare foarte atentă a testului din punctul de vedere al validităţii, fidelităţii şi adecvării la lotul de subiecţi, ceea ce nu se prea întâmplă în cazul testelor de cunoştinţe informative.

Cota brută a unui elev la un test de cunoştinţe elaborat de profesor aduce mai multe informaţii asupra individului respectiv dacă este comparată cu rezultatele altor persoane din aceeaşi clasă. Interpretând scorul unui elev prin raportare la poziţia sa în cadrul clasei se elimină erorile generate de neadecvarea testului la nivelul de cunoştinţe al elevilor examinaţi (un test prea dificil pentru o clasă furnizează doar scoruri mici, astfel încât cel mai bun rezultat obţinut poate fi situat sub ceea ce profesorul consideră că ar fi "nivelul minim necesar de performanţă").

13.3. INTERPRETAREA SCORURILOR LA TESTELE CENTRATE PE CRITERIU

Aceste teste folosesc, pentru interpretarea scorurilor, un *criteriu de performanţă* fixat înainte de utilizarea lor şi care este independent de rezultatele pe care le obţin la test diverse populaţii.

Stabilirea criteriului se face de către grupul de specialişti care a elaborat testul, ţinând cont de comportamentele aşteptate din partea persoanelor examinate şi de nivelul la care este necesar să se situeze rezultatele, pentru un individ sau pentru un grup de persoane.

De exemplu, în cazul unor teste de cunoştinţe, criteriile pot fi de următoarea formă: "Elevul trebuie să traducă cel puţin 80% din cuvintele cuprinse în test." sau "90% din grup trebuie să rezolve corect cel puţin 6 probleme din 10."

Utilizarea testelor centrate pe criteriu în programele instructiv-educative are ca principal scop clasificarea elevilor pe "niveluri" de stăpânire a cunoştinţelor, pentru fiecare din obiectivele urmărite de test (Swaminathan, Hambleton, Algina, 1974). Pe baza rezultatelor obţinute la test se decide dacă elevul poate trece la programul de instruire următor ori rămâne la programul actual, temporar sau un timp nedefinit.

Deci, pentru un test centrat pe criteriu sunt fixate de către specialişti două sau mai multe categorii ordonate, iar scorurile obţinute la test servesc la repartizarea subiecţilor în aceste categorii. Problema care trebuie rezolvată în acest caz constă în definirea categoriilor.

Să considerăm un test centrat pe criteriu, având m scoruri posibile: $s_1<s_2<...<s_m$. Se intenţionează ca acesta să fie folosit la clasificarea subiecţilor în k categorii ordonate ($2\leq k<m$), prima categorie corespunzând celor mai mici scoruri, iar ultima, celor mai mari.

Definirea categoriilor se face prin fixarea a k+1 praguri:

$$t_0 < t_1 <...< t_k \qquad (13.3.1)$$

care realizează o împărţire a intervalului $[s_1, s_m]$ în k intervale disjuncte. Pentru ca orice scor de test să îşi poată găsi locul într-o categorie, se impun următoarele condiţii:

- primul prag să fie mai mic decât cel mai mic scor posibil, iar următorul prag să fie mai mare decât acest scor:

$$t_0 < s_1 < t_1 \qquad (13.3.2)$$

- ultimul prag să coincidă cu cel mai mare scor posibil:

$$t_k = s_m \qquad (13.3.3)$$

În cea de-a j-a categorie (j=1,2,...,k) vor fi repartizate toate persoanele al căror scor aparţine intervalului $(t_{j-1},t_j]$.

Pragurile $t_0, t_1,..., t_k$ pot fi fixate de specialişti sau pot fi determinate astfel încât clasificarea definită cu ajutorul lor să semene cât mai mult cu o clasificare existentă, în acelaşi număr de clase, realizată de specialişti. Pentru cea de-a doua variantă, se poate utiliza următorul algoritm:

(1) Se administrează testul unei populaţii repartizate deja în k categorii. De exemplu, în cazul k=2, prima categorie poate cuprinde persoane apreciate de specialişti ca având un nivel de cunoştinţe "nesatisfăcător", iar a doua categorie, persoane cu un nivel "satisfăcător" al cunoştinţelor.

(2) Pentru fiecare posibilitate de fixare a pragurilor $t_0, t_1,...,t_k$, care îndeplineşte condiţiile (13.3.1), (13.3.2) şi (13.3.3):

- Se face clasificarea persoanelor, pe baza scorurilor obţinute la test, astfel: un individ care are scorul $x \in (t_{j-1}, t_j]$ este repartizat în categoria j, $j=1,2,...,k$.
- Se calculează coeficientul lui Cohen, K (a se vedea Anexa I), care evaluează concordanţa dintre clasificarea existentă şi cea corespunzătoare pragurilor $t_0, t_1,..., t_k$. Coeficientul K ia valori între -1 şi +1, fiind cu atât mai apropiat de +1, cu cât cele două clasificări se aseamănă mai mult între ele. Pentru fiecare alegere a pragurilor $t_0, t_1,..., t_k$ se obţine o altă valoare a coeficientului K.

(3) Se aleg drept margini ale celor k categorii pragurile $t_0, t_1,..., t_k$ pentru care coeficientul K este maxim. Acestea realizează o clasificare care se aseamănă cel mai mult posibil cu cea existentă deja în populaţie.

În cazul în care testul împarte persoanele examinate în două clase (k=2), pragul t_1 se numeşte *nivel acceptabil de performanţă* (Guilbert, 1987). Se consideră că un subiect îndeplineşte cerinţele criteriului dacă a realizat un scor mai mare decât t_1.

În literatura de specialitate există mai multe păreri privind modul de stabilire a nivelului acceptabil de performanţă. Se poate aplica algoritmul descris mai sus sau se poate alege drept barem pragul t_1 pentru care frecvenţa relativă a persoanelor repartizate în aceeaşi categorie în ambele clasificări, p_0, este maximă (Crehan, 1974). Pentru calculul frecvenţei p_0 se utilizează formula:

$$p_0 = \frac{\sum_{i=1}^{k} n_{ii}}{N} \qquad (13.3.4)$$

unde:

n_{ii} = numărul de persoane aflate în categoria i şi repartizate pe baza scorului obţinut la test tot în categoria i, i=1,2,...,k;

N = numărul total de persoane cărora li s-a administrat testul.

13.4. INTERPRETAREA SCORURILOR LA TESTELE NORMATIVE

13.4.1. Normele testelor

În cazul testelor normative, interpretarea scorurilor individuale se realizează cu ajutorul unor scări valorice, denumite **norme**, formate în cadrul operaţiei de etalonare a testului.

Prin **etalonare** se înţelege stabilirea unei scări care să permită determinarea locului ocupat de rezultatele unui subiect faţă de rezultatele unei populaţii de referinţă, suficient de numeroasă, formată din persoane comparabile cu cea examinată.

Obiectivul principal al etalonării îl constituie determinarea repartiţiei de frecvenţe a scorurilor testului într-un eşantion extras din această populaţie.

Normele testului descriu performanţa realizată de eşantionul de indivizi ales. Ele sunt valori ale unei anumite caracteristici a repartiţiei scorurilor în eşantion. Caracteristica urmărită poate fi:

- frecvenţa scorurilor mai mici sau mai mari decât o valoare dată

sau

- valoarea care împarte şirul scorurilor ordonate în două intervale ale căror frecvenţe sunt într-un raport fixat

sau

- media scorurilor.

Cu ajutorul normelor se determină modul în care se transformă cotele brute ale testului în cote uşor de interpretat, numite **cote transformate**. Acestea permit compararea rezultatelor între persoane diferite sau, pentru aceeaşi persoană, între teste diferite.

Pentru a fi cu adevărat utile, normele trebuie să fie descrise detaliat în manualul testului, indicând:

a. compoziţia populaţiei pentru care au fost stabilite;
b. modul de formare a eşantionului din această populaţie;
c. numărul persoanelor din eşantion;

d. caracteristicile eşantionului: sex, vârstă, rasă, statut socio-economic, localizare geografică, nivel de şcolarizare etc.

e. gradul în care eşantionul ales este reprezentativ pentru populaţia specificată;

f. natura şi uniformitatea condiţiilor în care a fost administrat testul;

g. data testării.

Tabelele de norme incluse în manualele testelor cuprind în paralel cotele brute şi cotele transformate corespunzătoare, pentru una sau mai multe populaţii. Ele oferă informaţii numai asupra rezultatelor "obişnuite" sau "tipice" pentru populaţiile respective. Nu este vorba despre rezultate "ideale", care ar trebui egalate.

Tabelele de norme ajută să se dea răspuns la următoarele întrebări:

- Cum este rezultatul la test al unei persoane în comparaţie cu cel al altor persoane din aceeaşi populaţie?
- Cum este rezultatul unei persoane la un test comparativ cu rezultatul la un alt test?
- Cum este rezultatul la un test comparativ cu rezultatul aceleiaşi persoane, la acelaşi test, administrat cu câtva timp înainte?

Pentru ca interpretarea unui scor individual să fie corectă trebuie ca normele la care se face raportarea acestuia să fie alese în conformitate cu scopul testării. De exemplu, subiectul va fi comparat cu un eşantion provenit din populaţia din care el face parte, dacă examinarea psihologică are scop diagnostic, sau cu un eşantion extras din populaţia din care ar urma să facă parte, dacă testul se foloseşte în cadrul selecţiei profesionale.

13.4.2. Etalonarea testelor

Etalonarea este una dintre etapele acţiunii de standardizare a testului. Ea se desfăşoară în următorii paşi:

(1) Se defineşte populaţia pentru care se etalonează testul.

(2) Se extrage un eşantion din această populaţie.

(3) Se administrează testul acestui eşantion.

(4) Se determină formula de obţinere a cotelor transformate din scorurile testului.

În continuare vom detalia conţinutul celor patru paşi.

(1) Când se construieşte un test, este important să se decidă cine sunt persoanele care vor fi examinate cu el, ce caracteristici are cel mai mare grup din care vor proveni ele. Acesta reprezintă *populaţia* căreia îi este destinat testul.

Definirea populaţiei se poate face fie concret, prin enumerarea tuturor membrilor ei, fie operaţional, prin indicarea unui set de reguli care vor permite să se cunoască, pentru orice persoană, dacă aparţine sau nu populaţiei.

De exemplu, populaţia poate fi formată din "toate persoanele care au vârsta cuprinsă între 14 ani împliniţi şi 18 ani neîmpliniţi, care locuiesc în oraşul Bucureşti şi sunt elevi la liceu".

Definiţia populaţiei pentru care se face etalonarea unui test este foarte arbitrară: toţi indivizii sau numai un lot cu acelaşi nivel şcolar sau de aceeaşi vârstă, dintr-un oraş, o regiune sau o ţară.

(2) Din populaţie se va extrage un eşantion de indivizi -numit *grup de normare*- căruia i se va administra testul, pentru a stabili normele.

Este important ca în eşantionul format testul să se comporte la fel ca în întreaga populaţie: să măsoare sau să prezică aceeaşi variabilă, iar repartiţia scorurilor la test în eşantion să fie asemănătoare celei care s-ar obţine dacă s-ar dispune de scorurile întregii populaţii.

Pentru aceasta, eşantionul constituit trebuie să îndeplinească două cerinţe:

- să fie reprezentativ pentru populaţia din care a fost extras (structura sa să fie asemănătoare structurii populaţiei, în privinţa zonelor geografice şi a mediului de domiciliu, a nivelului socio-economic, a vârstei şi a altor caracteristici pe care autorul testului le consideră importante şi care ar putea influenţa rezultatele testului);
- să fie destul de mare, aşa încât rezultatele obţinute la test să reflecte repartiţia acestor valori în populaţia din care provine.

Obţinerea unui eşantion se poate face prin selecţie aleatoare simplă sau prin selecţie aleatoare stratificată.

În primul caz, fiecare membru al populaţiei are şanse egale de a intra în componenţa eşantionului. Dacă eşantionul este numeros, atunci în el se vor găsi persoane aparţinând tuturor zonelor geografice, claselor sociale, naţionalităţilor prezente în populaţie, cu aceeaşi frecvenţă relativă ca în populaţie, deci eşantionul va fi reprezentativ pentru populaţie.

Atunci când volumul eşantionului este mic, pentru obţinerea unui eşantion reprezentativ este necesar să se facă o selecţie aleatoare stratificată. În acest scop, se determină variabilele de mediu (sex, vârstă, rasă, clasă socială, nivel educaţional etc.) care au o influenţă mare asupra rezultatelor aplicării testului. Se face apoi o clasificare a populaţiei după valorile acestor variabile şi se stabileşte care trebuie să fie compoziţia eşantionului astfel încât fiecare clasă să fie reprezentată în eşantion proporţional cu frecvenţa ei relativă în populaţie.

Exemplu.

Dacă populaţia este formată din 100000 de persoane şi se formează clase în funcţie de sex şi mediul de domiciliu, conform tabelului următor:

Clasa	Sex	Mediu de domiciliu	Frecvenţa	Frecvenţa relativă
C1	bărbaţi	urban	27000	0.27
C2	bărbaţi	rural	22000	0.22
C3	femei	urban	23000	0.23
C4	femei	rural	28000	0.28
Total			100000	1.00

atunci într-un eşantion compus din 2000 de persoane frecvenţele claselor vor fi:

Clasa	Frecvenţa relativă	Frecvenţa
C1	0.27	540
C2	0.22	440
C3	0.23	460
C4	0.28	560
Total	1.00	2000

Din fiecare clasă se extrag, prin selecţie aleatoare simplă, persoanele care vor intra în componenţa eşantionului.

Principala dificultate în cazul selecţiei aleatoare stratificate constă în deteminarea variabilelor după care se face împărţirea în clase. Se recomandă (Kline, 1993) ca ele să se aleagă dintre variabilele care corelează semnificativ cu scorurile testului. Numărul lor nu va fi prea mare, pentru a nu se forma prea multe clase, ceea ce ar necesita alcătuirea unui eşantion foarte numeros. În general nu se aleg mai mult de patru variabile. Frecvent, variabilele după care se face clasificarea sunt sexul, vârsta şi clasa socială.

Volumul eşantionului se stabileşte astfel încât din fiecare clasă să se extragă cel puţin 300 de persoane (Kline, 1993). Se deduce de aici că populaţiile generale, care sunt eterogene şi din acest motiv necesită utilizarea mai multor variabile pentru formarea claselor, impun obţinerea unor eşantioane foarte numeroase. Pentru a se face economie de timp şi de bani, adesea se renunţă la

stabilirea normelor pentru o populaţie generală şi se utilizează populaţii mai puţin numeroase şi mai omogene, pentru care eşantioanele pot fi de volum mai redus. În acest fel se obţin *norme locale ale testului.*

Dacă se constată că un subgrup al eşantionului, care poate fi identificat printr-o valoare a unei variabile de mediu (de exemplu "femei" sau "persoane din mediul urban") realizează la test scoruri mult mai mari sau mult mai mici decât restul eşantionului, se vor determina *norme ale subgrupului* şi acestea vor fi folosite în practică atunci când persoanele al căror scor trebuie interpretat au aceeaşi valoare a variabilei de mediu ca şi subgrupul.

(3) Dacă eşantionul are un volum foarte mare ori este format din persoane care domiciliază sau lucrează în locuri diferite, testul nu poate fi administrat deodată tuturor indivizilor. Examinarea cu ajutorul testului se va face pe subgrupe. Dar, pentru ca rezultatele să nu fie distorsionate, condiţiile de administrare trebuie să fie identice pentru toate subgrupele.

(4) În urma analizei scorurilor realizate la test de persoanele din eşantion se stabilesc normele testului. Apoi se determină modul în care se obţin cotele transformate din cotele brute. În acest scop se folosesc mai frecvent două tipuri de transformări:

• *transformări liniare*, care schimbă doar media şi abaterea standard a rezultatelor, păstrând nemodificată repartiţia cotelor, deci ordinea indivizilor din eşantion. Prin astfel de transformări se obţin **cotele standard *z*** şi **cotele standardizate;**

• *transformări neliniare*, sau *transformări de arie*, care schimbă repartiţia cotelor brute, făcând-o să semene cu repartiţia uniformă (în cazul **centilelor, decilelor** sau al **cuartilelor**) ori cu cea normală (în cazul când se urmăreşte obţinerea unei **scale normalizate** cu 5, 7 sau 9 trepte). Acest fel de transformări schimbă unitatea scalei, astfel încât frecvenţa relativă a cotelor brute care ocupă un interval de valori particular va fi egală cu aria suprafeţei cuprinsă între intervalul transformat, desenat pe axa absciselor, şi:

• curba repartiţiei uniforme, pentru rangurile centile, decile şi cuartile sau

• curba repartiţiei normale, pentru clasele scării normalizate.

Cotele transformate rezultate prin transformări liniare sau neliniare indică poziţia relativă a subiectului într-o populaţie.

Un alt tip de transformări utilizate în etalonare sunt cele care conduc la **clase echivalente** sau la **vârste echivalente**. Ele se aplică pentru teste care măsoară variabile fiziologice sau psihologice ale căror valori cresc cu vârsta (de exemplu înălţimea, greutatea, bogăţia vocabularului, performanţa la citire etc.).

Aceste cote transformate descriu performanţa subiectului sub forma indicării unei populaţii a cărei mediană sau medie a cotelor brute este aproximativ egală cu cota brută a persoanei respective. Unele teste de inteligenţă utilizează pentru interpretarea scorurilor, în locul vârstelor echivalente, normele de vârstă mintală.

13.4.3. Clasele echivalente

Aceste cote transformate sunt folosite frecvent pentru testele de cunoştinţe standardizate, destinate claselor primare.

Tabelele cu clasele echivalente conţin, pentru fiecare lună de şcoală a fiecărei clase, mediana cotelor brute la test. Clasa echivalentă este exprimată printr-un număr zecimal: partea întreagă reprezintă clasa, iar cifra zecimală, luna de şcoală. Se face presupunerea că în cele două luni de vacanţă de vară cotele la test ale elevilor nu se modifică.

Pentru interpretarea rezultatului obţinut de un subiect se caută în tabelul de norme clasa echivalentă corespunzătoare cotei brute respective.

Exemplu.

În tabelul 13.4.3.1. sunt notate clasele echivalente pentru un test de aritmetică ale cărui cote brute iau valori între 0 şi 100, destinat elevilor din clasele a II-a, a III-a şi a IV-a.

Tabelul 13.4.3.1

Clasele echivalente pentru un test de aritmetică

Clasa	Cota brută	Clasa	Cota brută	Clasa	Cota brută
2.0	15	3.0	25	4.0	55
2.1	15	3.1	27	4.1	58
2.2	16	3.2	28	4.2	62
2.3	17	3.3	30	4.3	64
2.4	17	3.4	32	4.4	66
2.5	18	3.5	36	4.5	67
2.6	19	3.6	39	4.6	68
2.7	20	3.7	44	4.7	69
2.8	22	3.8	47	4.8	69
2.9	23	3.9	51	4.9	70

Un elev care în prima săptămână de şcoală din clasa a IV-a a obţinut la test o cotă brută egală cu 27 are clasa echivalentă 3.1, deci este rămas în urmă la aritmetică cu aproape un an.

Popularitatea utilizării claselor echivalente este dată de faptul că furnizează cote transformate uşor de înţeles şi permite compararea rezultatelor obţinute la mai multe teste.

Exemplu.

Un copil, care este în a doua lună de şcoală din clasa a patra, a obţinut la trei teste următoarele clase echivalente:

Limba română : 4.5
Aritmetică : 3.0
Istorie : 5.4

Analizând aceste rezultate, părinţii şi profesorii vor înţelege că performanţele la Limba română sunt apropiate de cele ale celorlalţi elevi în această perioadă a anului şcolar, în timp ce scorul de la testul de Aritmetică denotă o rămânere în urmă de peste un an, iar cel de la testul de Istorie indică o depăşire a cunoştinţelor cu peste un an.

Trebuie subliniat că valoarea cotei brute indicată în tabelul de norme pentru fiecare clasă echivalentă nu reprezintă o performanţă care este necesar să fie atinsă. Ea este doar un rezultat tipic al unui grup de elevi cu acelaşi nivel de şcolarizare. Aproximativ 50% dintre persoanele grupului respectiv au rezultate mai bune şi aproximativ 50% au rezultate mai slabe decât acestă cotă.

Principala deficienţă a claselor echivalente constă în faptul că aceleiaşi "distanţe" între clase echivalente îi pot corespunde distanţe diferite între cotele brute ale testului. De exemplu, în tabelul 13.4.3.1, unei diferenţe de un an îi corespunde o distanţă de 10 puncte, dacă se compară clasele echivalente 2.0 şi 3.0, sau de 30 de puncte, dacă se compară clasele echivalente 3.0 şi 4.0. Deci, informaţia "Elevul este în urmă cu un an la aritmetică" are semnificaţie diferită, în funcţie de momentul în care este emisă.

Explicaţia constă în faptul că dezvoltarea intelectuală şi acumularea de cunoştinţe sunt procese care nu au o evoluţie liniară în timp. În perioadele şcolare în care acumularea de cunoştinţe este masivă, unei diferenţe mici de clase echivalente (una sau două luni şcolare) îi corespunde o diferenţă mare de cote brute la test. În plus, curba dezvoltării intelectuale relative la vârstă diferă de la o persoană la alta.

Din aceste motive se recomandă să se folosească transformarea cotelor brute în clase echivalente numai pentru aritmetică, citire şi compunere şi doar

în perioada claselor primare, când procesul de instruire este asemănător în toate şcolile.

Un procedeu care poate fi aplicat pentru obţinerea tabelelor cu clase echivalente este următorul (Flanagan, 1955):

(1) Se administrează testul, în acelaşi timp, unor loturi mari de elevi din clase consecutive, deci cu "distanţă" de o clasă echivalentă între ele.

(2) Pentru fiecare clasă se determină mediana cotelor brute.

(3) Se interpolează punctele consecutive (clasă, mediană), fie liniar, fie printr-o funcţie crescătoare.

(4) Pentru fiecare clasă echivalentă aflată între clasele pentru care s-a determinat mediana, din lună în lună, se citeşte de pe graficul funcţiei interpolatoare cota brută a testului şi se rotunjeşte la cel mai apropiat întreg.

Este important ca loturile de elevi utilizate la stabilirea claselor echivalente să fie numeroase şi să conţină persoane cu niveluri şcolare diferite (atât avansaţi, cât şi retardaţi).

13.4.4. Vârstele echivalente

Aceste cote transformate indică, pentru fiecare vârstă cronologică exprimată sub forma unei perechi de numere (an, lună), mediana sau media (după preferinţa celui care a făcut etalonarea) cotelor brute la test. Ele sunt folosite mai ales pentru teste de inteligenţă (tabelele de norme conţin "vârsta mintală" echivalentă) şi pentru teste de cunoştinţe (tabelele de norme conţin "vârsta la aritmetică" sau "vârsta la citire" etc.). La fel ca şi clasele echivalente, aceste cote transformate sunt potrivite pentru nivelul şcolar elementar, când dezvoltarea mintală şi influenţa educaţiei sunt continue şi nu oscilează.

Vârstele echivalente au aceleaşi avantaje şi deficienţe ca şi clasele echivalente.

13.4.5. Normele de vârstă mintală

Conceptul de *vârstă mintală* a fost introdus în scalele Binet-Simon revizuite în anul 1908. Acestea grupează itemii pe niveluri de vârstă: itemii care sunt rezolvaţi corect de majoritatea copiilor de 7 ani din eşantionul utilizat la etalonare sunt plasaţi la nivelul vârstei de 7 ani ş.a.m.d. Scorul unui copil la acest test corespunde celui mai înalt nivel pe care l-a rezolvat cu succes. De pildă, dacă un copil de 10 ani (vârsta cronologică) rezolvă corect itemii aflaţi la nivelul vârstei de 12 ani, se va spune că are vârsta mintală de 12 ani.

Trebuie menţionat că o unitate de vârstă mintală (un an mintal) nu are aceeaşi semnificaţie la orice vârstă cronologică. Datorită faptului că dezvoltarea intelectuală este rapidă la vârste fragede şi îşi încetineşte ritmul la maturitate, importanţa unei diferenţe de un an mintal scade cu vârsta. O diferenţă de un an mintal la vârsta cronologică de 3-4 ani înseamnă tot atât de mult ca o diferenţă de 3 ani mintali la vârsta cronologică de 9-12 ani (Anastasi, 1954).

Pentru a putea interpreta mai uşor rezultatele testelor de inteligenţă, fără a mai face referire de fiecare dată la vârsta cronologică, s-a definit *coeficientul de inteligenţă* (IQ = Intelligence Quotient). Acesta este egal cu raportul dintre vârsta mintală şi vârsta cronologică, multiplicat cu 100. Deci, o valoare de 100 a acestui coeficient indică egalitatea vârstei mintale cu vârsta cronologică.

Coeficientul de inteligenţă permite compararea persoanelor de vârste diferite în privinţa dezvoltării intelectuale. Dar, el nu poate fi utilizat pentru orice test de inteligenţă destinat adulţilor. La un test ca Stanford-Binet, de exemplu, media scorurilor obţinute de persoane de aceeaşi vârstă cronologică, mai mare de 15 ani, nu creşte semnificativ de la un an la altul. Deci, indiferent de vârsta lor cronologică, adulţii care nu sunt debili mintal vor obţine la test, în general, vârsta mintală de 15 ani. Numărătorul coeficientului de inteligenţă fiind acelaşi (egal cu 15), valoarea raportului prin care se calculează acesta va depinde doar de vârsta cronologică, mai precis va descreşte cu vârsta.

13.4.6. Cotele standard *z* şi cotele standardizate

Aceste cote transformate servesc la indicarea poziţiei unei persoane faţă de o anumită populaţie.

O **cotă standard *z*** arată cu câte unităţi de abatere standard se distanţează cota brută la test a unui subiect de media cotelor brute în populaţia cu care se face comparaţia.

Cotele *z* se calculează pe baza cotelor brute ale testului, folosind formula:

$$z = \frac{x-m}{\sigma} \qquad (13.4.6.1)$$

unde:
z = cota z corespunzătoare scorului la test x;
m, σ = media şi abaterea standard pentru cotele brute la test în eşantionul folosit la etalonare.

Cotele standard se folosesc pentru a compara performanţele unui individ la două sau mai multe teste, dacă mediile şi dispersiile scorurilor acestora au fost calculate pe baza aceluiaşi eşantion de persoane.

De exemplu, dacă un elev are cotele z la un test de aritmetică şi la unul de limba română egale cu 1.8, respectiv cu 0.3, se poate aprecia că elevul este mai bine pregătit la aritmetică decât la limba română, chiar dacă el a obţinut aceeaşi cotă brută la ambele teste.

Cotele standard sunt uşor de interpretat atunci când cotele brute ale testului sunt repartizate aproximativ normal în populaţia care a servit la etalonare. În acest caz, în intervalul [-1,+1] se află cotele z pentru 68% dintre persoane. Despre un subiect care are o cotă standard z=0.72 se poate afirma că a obţinut un rezultat "obişnuit", în timp ce o persoană cu o cotă z egală cu 3.6 are un scor "neobişnuit de mare" (mai puţin de 0.13% dintre cei testaţi cu ocazia etalonării au mai realizat un asemenea scor).

Cotele standardizate au aceleaşi proprietăţi ca şi cotele standard z, singura diferenţă între ele fiind valoarea mediei şi cea a abaterii standard (a se vedea Anexa II).

13.4.7. Rangurile centile, decile şi cuartile

Rangul centil arată câte persoane s-ar afla în urma unui anumit subiect, dacă acesta ar face parte dintr-un grup cu 100 de indivizi. De exemplu, dacă 85% din cei testaţi obţin scoruri mai mici decît 60 şi 15% realizează scoruri mai mari, atunci scorul 60 va avea rangul centil 85.

Prezentăm în continuare, pe un exemplu, paşii algoritmului de determinare a rangurilor centile, prelucrând rezultatele obţinute la un test de un eşantion format din 200 de persoane.

(1) Se ordonează crescător cotele brute ale testului (x) şi se notează pentru fiecare valoare numărul de persoane care au realizat scorul respectiv (N). Se obţine tabelul 13.4.7.1.

Tabelul 13.4.7.1
Frecvenţele cotelor brute la test

x	10	11	12	14	15	16	18	21	22	23
N	10	20	20	30	10	40	30	20	10	10

(2) Se trasează poligonul frecvenţelor cumulate (Figura 13.4.7.1), procedînd astfel:

- În jurul fiecărei valori de scor x_i, i=1,2,...,n, se construieşte intervalul [x_i-0.5, x_i+0.5].
- Se calculează valoarea fiecărei ordonate y_{i-}, i=1,2,...,n, corespunzătoare abscisei x_i-0.5, făcând raportul dintre numărul persoanelor care au scor mai mic decît x_i şi numărul total de persoane în eşantion şi apoi înmulţind cu 100.
- Se calculează valoarea fiecărei ordonate y_{i+}, i=1,2,...,n, corespunzătoare abscisei x_i+0.5, făcând raportul dintre numărul persoanelor care au scor mai mic sau egal cu x_i şi numărul total de persoane în eşantion şi apoi înmulţind cu 100. Se observă că y_{i+} este egal cu $y_{(i+1)-}$.
- Se unesc, pe rând, punctele (x_1-0.5, y_{1-}), (x_1+0.5,y_{1+}), (x_2-0.5, y_{2-}),..., (x_n+0.5, y_{n+}), prin segmente de dreaptă.
- În exemplul nostru n=10. Valorile x_i, y_{i-} şi y_{i+}, i=1,2,...,10, sunt notate în tabelul 13.4.7.2.

Tabelul 13.4.7.2

Valorile x_i, y_{i-} şi y_{i+} utilizate la trasarea poligonului frecvenţelor cumulate

i	1	2	3	4	5	6	7	8	9	10
x_i	10	11	12	14	15	16	18	21	22	23
y_{i-}	0	5	15	25	40	45	65	80	90	95
y_{i+}	5	15	25	40	45	65	80	90	95	100

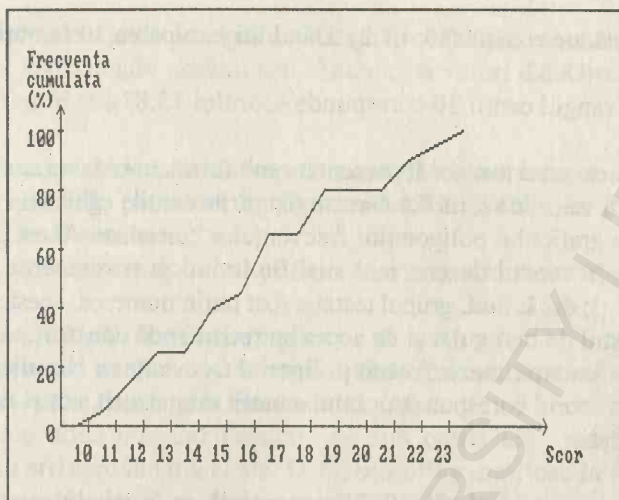


Figura 13.4.7.1
Poligonul frecvenţelor cumulate

(3) Pentru a determina rangul centil corespunzător unui scor dat, x, se citeşte de pe grafic ordonata punctului care are abscisa x. De exemplu, pentru scorul x_i=14:

$$i=4,\ y_{i-}=25,\ y_{i+}=40.$$

Rangul centil este valoarea y care corespunde scorului x=14 în ecuaţia dreptei ce trece prin punctele

$$(x_i-0.5, y_{i-})\ \text{şi}\ (x_i+0.5, y_{i+}):$$

$$\frac{y-25}{40-25} = \frac{x-13.5}{14.5-13.5}$$

sau

$$y = 15x - 177.5$$

Deci, scorul 14 are rangul 15*14-177.5 = 32.5.

Obişnuit, rangul centil se rotunjeşte la cel mai apropiat întreg.

(4) Pentru a determina scorul care corespunde unui rang centil dat, y, se citeşte de pe grafic abscisa punctului care are ordonata y.

De exemplu, rangului centil 30 îi corespunde pe grafic un punct situat pe dreapta care trece prin punctele (13.5 , 25) şi (13.5 , 40).

Această dreaptă are ecuaţia (13.4.7.2). Dând lui y valoarea 30 se obţine: $x = (30 + 177.5) / 15 = 13.83$

Deci rangul centil 30 corespunde scorului 13.83.

Normele unui test pot fi prezentate sub forma unui tabel care să conţină centilele, adică valorile scorurilor care au rangurile centile egale cu 1, 2, ..., 100, sau sub forma graficului poligonului frecvenţelor cumulate. Acest grafic poate fi realizat fie în modul descris mai sus, fie unind prin segmente de dreaptă punctele (x_i, y_{i+}); dacă, însă, grupul testat a fost puţin numeros, aceste puncte pot fi dispuse destul de neregulat şi de aceea se recomandă construirea unei curbe netede care să le aproximeze. Având poligonul frecvenţelor cumulate, se poate determina atât scorul corespunzător unui anumit rang centil, cât şi rangul centil al unui scor dat.

Unii autori (Lindvall, 1967) recomandă ca în tabelul de norme să se treacă pentru fiecare cotă brută a testului, în locul rangului centil, "banda centilă" care corespunde intervalului de încredere al scorului real. Pentru aceasta, după trasarea poligonului frecvenţelor cumulate, pentru fiecare scor posibil la test, x, se procedează astfel:

(1) se determină intervalul de încredere al scorului real, (x_-, x_+);

(2) se citesc de pe graficul poligonului frecvenţelor cumulate ordonatele corespunzătoare absciselor x_- şi x_+, y_- şi, respectiv, y_+;

(3) în tabelul de norme, în dreptul cotei brute x se notează banda centilă (y_-, y_+).

Tabelul 13.4.7.3 conţine un fragment dintr-un tabel cu benzi centile.

Tabelul 13.4.7.3

Exemplu de tabel cu benzi centile

Cota brută	Intervalul de încredere pentru scorul real	Banda centilă
...	...	...
15	[14, 16]	[5, 8]
16	[14, 17]	[5, 10]
17	[16, 18]	[8, 12]
...	...	...

Normele centile prezintă deficienţe în cazul când scorurile testului sunt repartizate aproximativ normal sau când există valori de scoruri care apar cu frecvenţe foarte mari.

Să considerăm, de exemplu, cazul când scorurile au o repartiţie de frecvenţe normală, prezentat în figura 13.4.7.2. Notăm cu f densitatea de repartiţie a legii normale.

Pentru oricare două valori de scor, $x_1<x_2$, aria suprafeţei delimitată de punctele

$(x_1,0), (x_2,0), (x_2,f(x_2)), (x_1,f(x_1))$

este proporţională cu frecvenţa relativă, în întregul eşantion, a persoanelor care au obţinut scoruri cuprinse în intervalul $(x_1,x_2]$.

Cele două porţiuni haşurate de sub curba legii normale, în figura 13.4.7.2, au arii aproximativ egale. Deci, scorurilor cuprinse în intervalul $(a,b]$ le va corespunde un interval al rangurilor centile având aproximativ aceeaşi lungime ca şi intervalul determinat de rangurile centile ale scorurilor cuprinse în intervalul $(c,d]$. Dar, intervalul de valori al scorurilor din prima zonă este de 4 ori mai mic decît cel al scorurilor din porţiunea a doua:

$d-c = 4\ (b-a)$

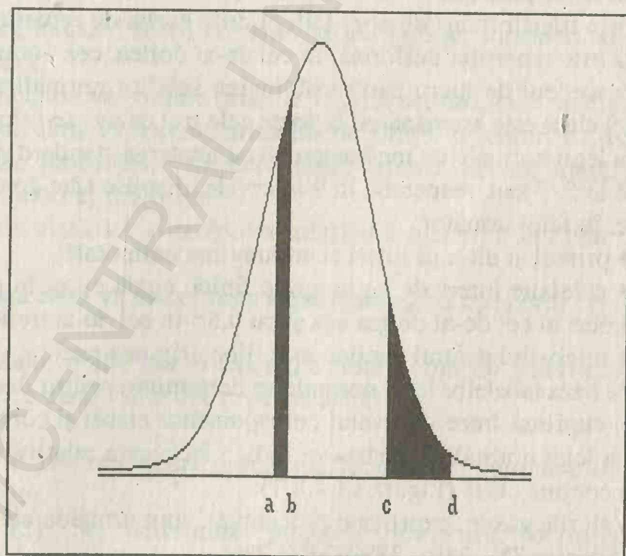


Figura 13.4.7.2
Două suprafeţe cu arii egale pentru o repartiţie normală

Înseamnă că două ranguri centile consecutive vor corespunde unor valori de scor foarte apropiate, deci unor persoane foarte asemănătoare -în privinţa variabilei măsurate de test-, în primul caz, şi unor scoruri mai îndepărtate, deci unor indivizi destul de deosebiţi unul de altul, în cel de-al doilea.

Atunci când repartiţia scorurilor testului este uniformă, deci fiecare cotă brută apare cu aceeaşi frecvenţă, unor intervale egale ca lungime pentru cotele brute le corespund intervale de lungimi egale pentru rangurile centile.

Folosirea rangurilor centile îşi găseşte justificarea numai atunci când testul are un număr mare de scoruri posibile. În caz contrar, este de preferat rangul decil sau rangul cuartil. Acestea se obţin într-un mod asemănător rangurilor centile, doar că aria cuprinsă între axa absciselor şi curba de repartiţie a scorurilor nu se mai împarte în 100 de părţi egale, ci în 10, pentru decile, sau în 4, pentru cuartile. Centilele care au rangurile 10, 20,..., 90 coincid cu decilele, iar centilele cu rangurile 25, 50 şi 75 sunt egale cu cuartilele.

13.4.8. Clasele scării normalizate

La fel ca şi rangurile centile, decile şi cuartile, clasele scării normalizate se obţin prin transformări de arie. Diferă doar curba de repartiţie utilizată: în primul caz este repartiţia uniformă, în cel de-al doilea, cea normală.

Procedeul de lucru pentru obţinerea scărilor normalizate cu 5, 7 şi, respectiv, 9 clase este asemănător. În toate cele trei cazuri se foloseşte curba de repartiţie a legii normale cu media zero şi cu abaterea standard σ. Axa reală se decupează în 5, 7 sau, respectiv, în 9 intervale, dispuse câte două simetric faţă de origine, în felul următor:

- primul şi ultimul interval au lungime nelimitată;
- celelalte intervale au lungime finită, egală cu σ, în primul caz, cu 0.66σ în cel de-al doilea caz şi cu 0.5σ în cel de-al treilea caz;
- intervalul central conţine în mijloc originea axei.

Pe baza tabelelor legii normale se determină, pentru fiecare clasă, aria suprafeţei cuprinsă între intervalul corespunzător clasei şi curba densităţii de repartiţie a legii normale. Aceasta va indica frecvenţa relativă a scorurilor pe care le va conţine clasa (Figura 13.4.8.1).

Valorile găsite, exprimate procentual, sunt următoarele :

- pentru 5 clase: 7%, 24%, 38%, 24%, 7%;
- pentru 7 clase: 5%, 11%, 21%, 26%, 21%, 11%, 5%;
- pentru 9 clase: 4%, 7%, 12%, 17%, 20%, 17%, 12%, 7%, 4%.

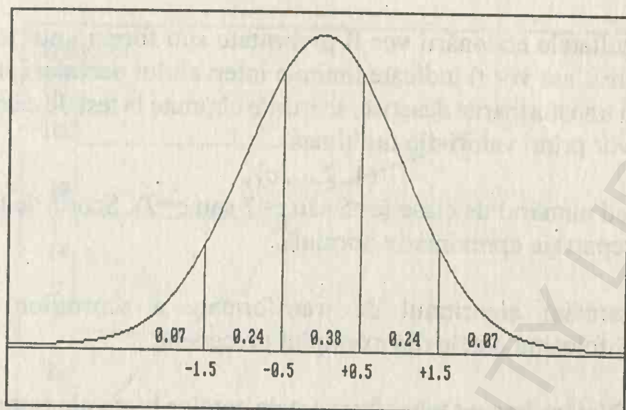


Figura. 13.4.8.1
Împărţirea suprafeţei cuprinsă sub curba densităţii de repartiţie a legii normale în cinci clase.

Concret, în cazul scării normalizate cu 5 clase, considerând scorurile ordonate crescător:

- prima clasă va conţine primele 7% din scoruri, cele mai mici;
- a doua clasă va conţine următoarele 24% din scoruri ş.a.m.d.

Pe baza procentelor se determină, pentru fiecare clasă, limitele intervalului de valori al scorurilor:

a. Prima clasă va avea limita inferioară egală cu cel mai mic scor posibil.

b. Ultima clasă va avea limita superioară egală cu scorul cel mai mare posibil.

c. Celelalte limite vor fi calculate ţinînd cont de valorile scorurilor cuprinse în fiecare clasă. Se poate proceda astfel :

(1) Se desenează poligonul frecvenţelor cumulate ale scorurilor.

(2) Se determină punctele graficului corespunzătoare frecvenţelor cumulate ale claselor (7%, 31%, 69%, 93% şi 100% pentru scara normalizată cu 5 clase).

(3) Se citesc de pe abscisă valorile scorurilor pentru aceste puncte; ele vor fi limitele intervalelor.

Rezultatele etalonării vor fi prezentate sub forma unui tabel, în care pentru fiecare clasă vor fi indicate limitele intervalului de valori al scorurilor.

Prin transformările descrise, scorurile obţinute la test de către eşantionul selecţionat vor primi valori din mulţimea

$$\{1, 2, ..., c\},$$

c reprezentînd numărul de clase (*c*=5 sau *c*=7 sau *c*=9). Scorurile transformate vor avea o repartiţie aproximativ normală.

Prezentăm algoritmul de transformare a scorurilor în 5 clase normalizate, folosind datele din exemplul precedent:

(1) Notăm într-un tabel frecvenţele cotelor brute ale testului (Tabelul 13.4.8.1).

Tabelul 13.4.8.1

Frecvenţele absolute şi cumulate ale scorurilor testului

Scor (*x*)	10	11	12	14	15	16	18	21	22	23
Frecvenţa absolută	10	20	20	30	10	40	30	20	10	10
Frecvenţa cumulată	10	30	50	80	90	130	160	180	190	200
Frecvenţa cumulată, în procente (*y*)	5	15	25	40	45	65	80	90	95	100

(2) Reprezentăm grafic frecvenţele cumulate, interpolând liniar între punctele (*x*,*y*) consecutive (Figura 13.4.8.2)

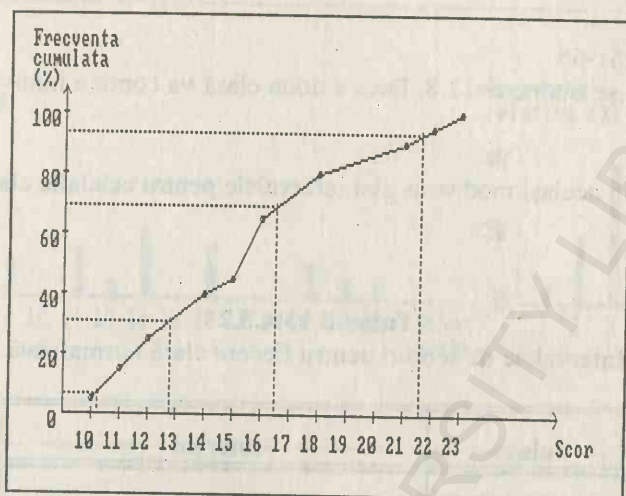


Figura 13.4.8.2

Determinarea intervalelor scorurilor la transformarea în cinci clase normalizate

(3) Prima clasă trebuie să cuprindă 7% din persoane. Punctul de pe grafic care are ordonata egală cu 7 se găseşte pe dreapta care trece prin punctele (10, 5) şi (11, 15). Ecuaţia acestei drepte este

$$\frac{y-5}{15-5} = \frac{x-10}{11-10}$$

sau

y=10x-95

Pentru y=7 se obţine x=10.2. Prima clasă va avea limita inferioară egală cu cea mai mică valoare posibilă de scor şi limita superioară egală cu 10.2.

(4) Primele două clase trebuie să cuprindă 7%+24%=31% din totalul persoanelor. Punctul de pe grafic care are ordonata egală cu 31 se află pe dreapta care trece prin punctele (12, 25) şi (14, 40). Această dreaptă are ecuaţia :

$$\frac{y-25}{40-25} = \frac{x-12}{14-12}$$

sau

$y=7.5x-65$

Pentru $y=31$ se obţine $x=12.8$. Deci a doua clasă va conţine intervalul (10.2 , 12.8].

(5) În acelaşi mod vom găsi intervalele pentru celelalte clase (Tabelul 13.4.8.2).

Tabelul 13.4.8.2

Intervalele de scoruri pentru fiecare clasă normalizată

clasa	interval
1	≤ 10.2
2	(10.2, 12.8]
3	(12.8, 16.5]
4	(16.5, 21.6]
5	> 21.6

(6) Utilizând tabelul 13.4.8.2 se determină clasa normalizată care corespunde fiecărei cote brute (Tabelul 13.4.8.3).

Tabelul 13.4.8.3

Corespondenţa cotă brută - clasă normalizată

cota brută	≤ 10	11	12	13	14	15	16	17	18	19	20	21	≥ 22
clasa normalizată	1	2	2	3	3	3	3	4	4	4	4	4	5

Prin această transformare se schimbă forma repartiţiei scorurilor (Figura 13.4.8.3).

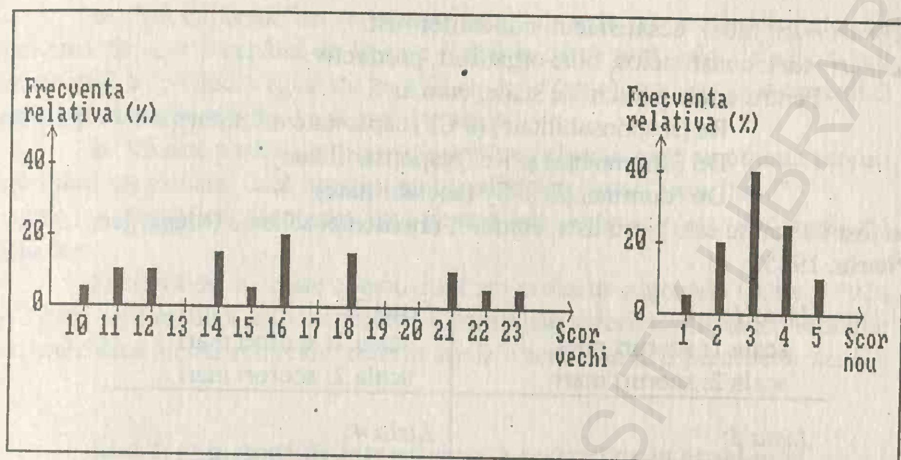


Figura 13.4.8.3

Schimbarea formei repartiţiei scorurilor prin transformarea în cinci clase normalizate

Există şi cazuri când o transformare în clase normalizate nu are sens. De exemplu, atunci cînd frecvenţa relativă a primului scor este mai mare decît 11% este imposibil de stabilit limitele claselor 1 şi 2 pentru o transformare normalizată în 9 clase.

13.5. INTERPRETAREA SCORURILOR LA CHESTIONARELE DE PERSONALITATE

13.5.1. Descrierea procedeului obişnuit

Interpretarea scorurilor obţinute de un subiect la majoritatea chestionarelor de personalitate are la bază adjective şi propoziţii scurte descriind comportamente, care sunt grupate în liste ataşate scalelor sau unor grupări de scale.

Prezentăm câteva exemple referitoare la Inventarul Psihologic California (CPI).

Pentru fiecare dintre primele 20 de scale din CPI există două asemenea liste: una corespunzătoare scorurilor mici şi alta corespunzătoare scorurilor mari.

De exemplu, scala So (care evaluează gradul de interiorizare a valorilor normative şi culturale) are ataşate următoarele liste (Pitariu, Albu, 1993):

Scoruri mici: rebel, nesatisfăcut, nonconformist.
Scoruri mari: conştiincios, bine organizat, productiv.

Pentru unele perechi de scale, cum ar fi

Re (responsabilitate) şi Cs (capacitate de statut),
Do (dominanţă) şi Re (responsabilitate),
Do (dominanţă) şi Sy (sociabilitate)

au fost întocmite câte patru liste, conform următoarei scheme (Megargee, 1972; Pitariu, 1985):

Lista 1: scala 1: scoruri mici scala 2: scoruri mari	*Lista 2*: scala 1: scoruri mari scala 2: scoruri mari
Lista 3: scala 1: scoruri mici scala 2: scoruri mici	*Lista 4*: scala 1: scoruri mari scala 2: scoruri mici

Pentru ca o caracterizare întocmită cu ajutorul listelor ataşate scalelor să concorde cât mai mult cu situaţia reală trebuie să fie soluţionate corect următoarele două probleme:

- Adjectivele sau propoziţiile conţinute în liste să fie adecvate scalei sau grupărilor de scale.
- Din liste să fie selectate acele adjective sau propoziţii care au o probabilitate mare de a fi potrivite persoanei testate.

Algoritmii utilizaţi pentru rezolvarea acestor probleme în cazul chestionarelor de personalitate utilizate în practică nu sunt totdeauna corecţi. Aşa este, de exemplu, metoda aplicată la formarea listelor ataşate scalelor din CPI (Gough, 1989, 1991; Megargee, 1972). Ea se bazează pe coeficientul de corelaţie liniară între scorurile scalelor şi scorurile adjectivelor (scorul unui subiect la un adjectiv este egal cu frecvenţa alegerii adjectivului pentru caracterizarea persoanei). În lista corespunzătoare scorurilor mari ale unei scale sunt introduse adjective care corelează pozitiv semnificativ statistic cu scorurile scalelor, deci care sunt mai potrivite persoanelor care obţin scoruri mari decât celor care realizează scoruri mici. Dar, nu există siguranţa că aceste adjective sunt într-adevăr potrivite pentru majoritatea celor care au scoruri mari şi, de asemenea, nu se poate şti dacă ele nu sunt valabile pentru un procent însemnat din mulţimea subiecţilor cu scoruri mici.

Faptul că listele sunt ataşate "scorurilor mari", respectiv "scorurilor mici", ale scalelor prezintă mai multe dezavantaje şi constituie o sursă de erori în interpretare. Astfel:

a. Prin expresia "scoruri mari (mici)" nu este precizată destul de clar valoarea de scor începând de la care (până la care) sunt valabile adjectivele listei; unii o consideră egală cu m, alţii cu $m+\sigma$ ($m-\sigma$), unde m şi σ reprezintă media şi abaterea standard a scorurilor scalei.

b. Nu este posibilă o diferenţiere între subiecţii care au obţinut scoruri apropiate ca valoare, dacă acestea sunt considerate "mari" sau "mici".

c. Nu există asemenea liste pentru interpretarea scorurilor medii ale scalelor.

Pornind de la aceste consideraţii am proiectat algoritmi (Albu, 1992a şi 1994) care rezolvă ambele probleme formulate anterior, utilizând "funcţiile caracteristice ale adjectivelor" referitoare la o scală sau la o pereche de scale.

13.5.2. O metodă de interpretare a scorurilor în cazul unei singure scale

Se consideră o scală a unui chestionar de personalitate, s, ale cărei scoruri pot lua valorile 0,1,...,n, şi un adjectiv (sau o expresie descriptivă) **a**.

DEFINIŢIA 13.5.2.1. (Albu, 1992). Prin **funcţie caracteristică** a adjectivului a relativă la scala s se înţelege o funcţie

$$f_c: \{0,1,...,n\} \to [0,1]$$

care ia ca valoare, în orice punct $x \in \{0,1,...,n\}$, probabilitatea ca adjectivul să fie potrivit pentru caracterizarea unei persoane care a obţinut la scala s scorul x.

Algoritmul de construire a funcţiei caracteristice a adjectivului **a** relativă la scala s se compune din doi paşi:

Pasul 1:

Se obţin punctele experimentale. Pentru aceasta se procedează astfel:

A. Se administrează scala s unui eşantion de subiecţi **Z**, cât mai numeros şi mai eterogen, astfel încât să se obţină cât mai multe valori de scor din mulţimea $\{0,1,...,n\}$.

B. Aceloraşi persoane li se cere să aprecieze dacă adjectivul **a** li se potriveşte sau nu.

C. Se notează:

N = numărul de persoane din lotul experimental, **Z**;

S = mulţimea scorurilor $x \in \{0,1,2,...,n\}$ care au fost obţinute la scala s de persoanele din lotul **Z**.

D. Se defineşte funcţia

$$f: \mathbf{S} \text{->} [0,1]$$

astfel:

$$f(x) = \frac{m(x)}{n(x)}, \quad x \in S \tag{13.5.2.1}$$

unde:

$n(x)$ = numărul de subiecţi care au obţinut la scala s scorul x:

$$\sum_{x \in S} n(x) = N$$

$m(x)$ = numărul de persoane, dintre cele care au obţinut scorul x la scală, care au considerat că adjectivul **a** li se potriveşte;

E. Se reprezintă grafic punctele experimentale $(x, f(x))$, $x \in$S.

Pasul 2:

Se construieşte funcţia caracteristică a adjectivului **a** relativă la scala s:

$$f_c: \{0,1,\ldots,n\} \text{->} [0,1]$$

al cărei grafic aproximează punctele experimentale.

Se procedează astfel:

A. Se determină, din mulţimea polinoamelor de grad cel mult trei, funcţia de cea mai bună aproximare, în sensul celor mai mici pătrate, a funcţiei f pe mulţimea S, notată cu P;

$P \in \{p: \mathbf{R} \text{->} \mathbf{R} \mid p(x) = ax^3 + bx^2 + cx + d,\ a,b,c,d \in \mathbf{R}\}$

B. Se defineşte funcţia

$$f_c: \{0,1,\ldots,n\} \text{->} [0,1]$$

astfel:

$$f_c(x) = \begin{cases} P(x), & \text{dacă } 0 \le P(x) \le 1 \\ 1, & \text{dacă } P(x) > 1 \\ 0, & \text{dacă } P(x) < 0 \end{cases} \tag{13.5.2.2}$$

pentru $x \in \{0,1,\ldots,n\}$.

C. Se calculează **eroarea medie de reprezentare**:

$$E = \frac{1}{N} \sum_{x \in S} n(x) |f(x) - f_c(x)| \tag{13.5.2.3}$$

DEFINIŢIA 13.5.2.2. Se consideră că adjectivul **a** este **adecvat** scalei **s**, la pragul de probabilitate $P_0 \in (0,1)$ fixat şi la pragul de eroare fixat ϵ dacă sunt îndeplinite următoarele condiţii:

- eroarea medie de reprezentare E este mai mică decât ϵ;
- funcţia caracteristică, f_c, îndeplineşte una dintre următoarele condiţii:

(C1) f_c este nedescrescătoare, adică:

$$t_1 < t_2 \Rightarrow f_c(t_1) \le f_c(t_2) \tag{13.5.2.4}$$

(C2) f_c este necrescătoare, adică:

$$t_1 < t_2 \Rightarrow f_c(t_1) \ge f_c(t_2) \tag{13.5.2.5}$$

(C3) există un punct $t_0 \in \{1,2,...,n-1\}$ astfel încât, oricare ar fi punctele $t_i \in \{0,1,...,n\}$, i=1,2,3,4, care îndeplinesc condiţia $t_1 \le t_2 \le t_0 \le t_3 \le t_4$, are loc una dintre următoarele situaţii:

$$f_c(t_1) \le f_c(t_2) \le f_c(t_0) \text{ şi } f_c(t_0) \ge f_c(t_3) \ge f_c(t_4) \tag{13.5.2.6}$$

sau

$$f_c(t_1) \ge f_c(t_2) \ge f_c(t_0) \text{ şi } f_c(t_0) \le f_c(t_3) \le f_c(t_4) \tag{13.5.2.7}$$

- funcţia f_c are pe mulţimea $\{0,1,...,n\}$ valori atât mai mari, cât şi mai mici decât P_0.

Toate adjectivele care sunt adecvate scalei s se reţin pentru a fi utilizate la caracterizarea persoanelor.

Tehnica de selecţie a adjectivelor pentru editarea raportului psihologic

Se consideră un chestionar de personalitate ale cărui scale sunt $s_1, s_2, ..., s_m$.

Se fixează un prag de probabilitate $P_0 \in (0,1)$ şi un prag de eroare ϵ.

Pentru fiecare scală s_i, $i=1,2,...,m$, se notează cu A mulţimea adjectivelor adecvate scalei s_i la pragul de probabilitate P_0 şi la pragul de eroare ϵ.

Pentru o persoană care la scalele chestionarului a obţinut scorurile $x_1, x_2, ..., x_m$ (x_i= scorul la scala s_i, $i=1,2,...,m$) editarea raportului psihologic se poate face astfel:

- la fiecare scală s_i, $i=1,2,...,m$, se determină valoarea funcţiei caracteristice a fiecărui adjectiv a din mulţimea A_i, în punctul x_i, $f_{c,a}(x_i)$;
- se reţin adjectivele pentru care $f_{c,a}(x_i) \ge P_0$.

13.5.3. O metodă de interpretare a scorurilor în cazul unei perechi de scale

Se consideră un adjectiv a şi două scale ale unui chestionar de personalitate, s_1 şi s_2.

Se notează cu X şi Y mulţimile scorurilor posibile la cele două scale.

DEFINIŢIA 13.5.3.1. (Albu, 1994). Prin **funcţie caracteristică** a adjectivului a relativă la perechea de scale (s_1,s_2) înţelegem o funcţie

$$F_c : XxY \to [0,1]$$

care ia ca valoare, în orice punct $(x,y) \in XxY$, probabilitatea ca adjectivul **a** să fie potrivit pentru caracterizarea persoanelor care au obţinut scorul x la scala s_1 şi scorul y la scala s_2.

Se notează:
N = numărul de persoane în lotul experimental;
$S \subseteq XxY$ = mulţimea perchilor de scoruri (x,y) obţinute la cele două scale de persoanele din lotul experimental;
$n(x,y)$ = numărul de subiecţi care au obţinut scorul x la scala s_1 şi scorul y la scala s_2, $(x,y) \in S$:

$$\sum_{(x,y)\in S} n(x,y) = N$$

$m(x,y)$ = numărul de subiecţi, dintre cei care au obţinut scorul x la scala s_1 şi scorul y la scala s_2, care au apreciat că adjectivul **a** li se potriveşte, $(x,y) \in S$.

Se defineşte funcţia:

$$F: S \to [0,1]$$

astfel:

$$F(x,y) = \frac{m(x,y)}{n(x,y)}, \quad (x,y) \in S \qquad (13.5.3.1)$$

Se determină funcţia de cea mai bună aproximare, în sensul celor mai mici pătrate, a funcţiei F, pe mulţimea S, notată cu P:
$P \in \{p: RxR \to R \mid p(x,y)=ax^3+bx^2+cx+dy^3+ey^2+fy+g,\ a,b,c,d,e,f,g \in R\}$

Se construieşte funcţia caracteristică a adjectivului **a**:

$$F_c : XxY \to [0,1]$$

astfel:

$$F_c(x,y) = \begin{cases} P(x,y), & \text{dacă } 0 \le P(x,y) \le 1 \\ 1, & \text{dacă } P(x,y) > 1 \\ 0, & \text{dacă } P(x,y) < 0 \end{cases} \qquad (13.5.3.2)$$

pentru $(x,y) \in$ XxY
şi se calculează **eroarea medie de reprezentare**, prin formula:

$$E = \frac{1}{N} \sum_{(x,y) \in S} n(x,y) \left| F(x,y) - F_c(x,y) \right| \qquad (13.5.3.3)$$

DEFINIŢIA 13.5.3.2. Se consideră că adjectivul **a** este **adecvat** perechii de scale (s_1,s_2), la pragul de probabilitate $P_0 \in (0,1)$ şi la pragul de eroare fixat ϵ, dacă funcţia sa caracteristică F_c, îndeplineşte următoarele condiţii:

- eroarea medie de reprezentare E este mai mică decât ϵ;
- restricţia funcţiei F_c pe orice submulţime a mulţimii **XxY** formată din puncte situate pe aceeaşi dreaptă este sau necrescătoare sau nedescrescătoare sau neconcavă sau neconvexă;
- funcţia F_c are pe mulţimea **XxY** valori atât mai mari cât şi mai mici decât P_0.

Toate adjectivele adecvate unei perechi de scale vor fi reţinute şi vor fi utilizate la descrirea persoanelor.

Pentru caracterizarea unui individ care a obţinut la o pereche de scale scorurile (x,y), vor fi selectate acele adjective ale căror funcţii caracteristice satisfac inegalitatea $F_c(x,y) \ge P_0$.

13.5.4. Avantajele celor două metode de interpretare a scorurilor

a. Metodele prezentate permit să se utilizeze, la caracterizarea persoanelor, nu numai adjective corespunzătoare valorilor extreme ale scorurilor scalelor, ci şi adjective care sunt potrivite doar scorurilor medii. De exemplu, pentru scala de flexibilitate Fx din CPI adjectivul "adaptabil" este potrivit numai scorurilor medii, la scoruri mari fiind mai potrivit adjectivul "instabil".

b. Ele fac posibilă caracterizarea prin adjective diferite a unor persoane care au scoruri apropiate ca valoare la una sau la două scale.

c. Pot fi utilizate pentru editarea rapoartelor psihologice cu ajutorul calculatorului.

13.5.5. Exemple

Un lot de 436 femei a răspuns la itemii din CPI.
S-au ales pragurile P_0=0.85 şi ϵ=0.20.

A. S-au construit funcţiile caracteristice ale propoziţiei:

(*p1*) "Nu are încredere în forţele proprii"

relative la fiecare dintre scalele Do (dominanţă) şi In (independenţă) din CPI, întrucât această propoziţie este inclusă în listele ataşate celor două scale, pentru caracterizarea persoanelor care obţin "scoruri mici".

Valorile statistice ale scorurilor celor două scale sunt :

- pentru Do: m=18.63 σ=5.76
- pentru In: m=15.37 σ=3.74

În figurile 13.5.5.1 şi 13.5.5.2 sunt reprezentate graficele restricţiilor funcţiilor caracteristice ale propoziţiei *p1* relative la aceste scale, pe mulţimea scorurilor obţinute de subiecţi.

S-a putut constata că scorul maxim pentru care propoziţia *p1* este potrivită cu o probabilitate mai mare decât P_0 este mai mic, în ambele cazuri, decât m-σ: pentru Do acest scor este egal cu 11, iar pentru In, este egal cu 10.

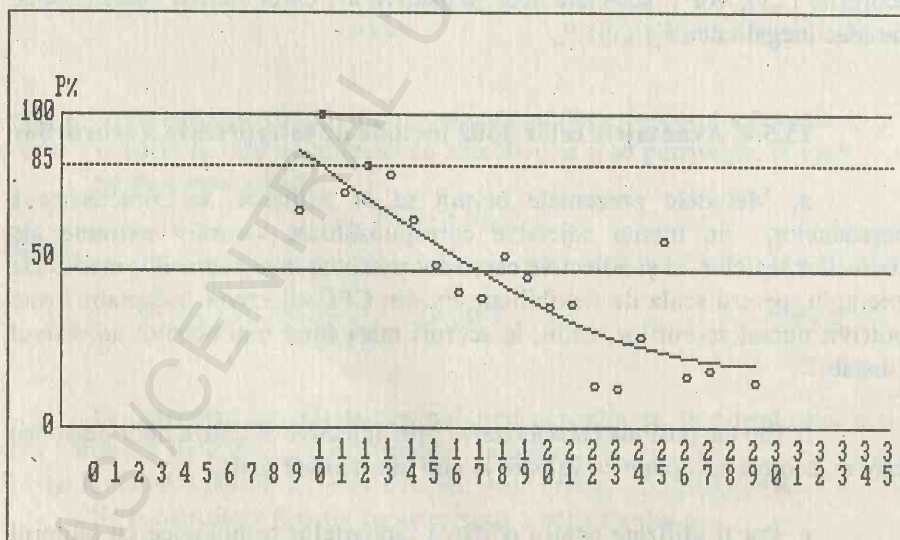


Figura 13.5.5.1
Graficul funcţiei caracteristice a propoziţiei *p1*, relativă la scala Do

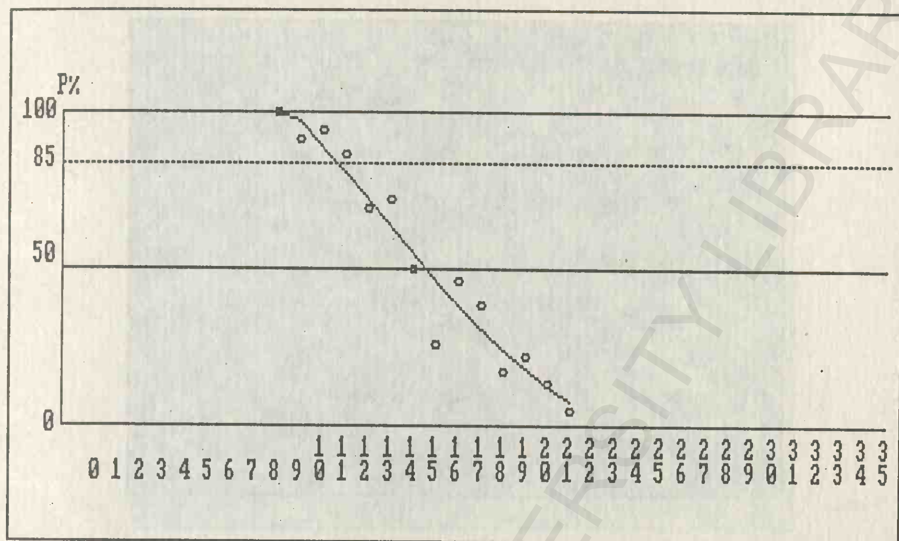


Figura 13.5.5.2
Graficul funcţiei caracteristice a propoziţiei *p1* relativă la scala In

B. S-au construit funcţiile F_c, pentru propoziţiile:
(*p2*) Simt o dorinţă puternică de reuşită în viaţă.
(*p3*) Mă entuziasmez foarte uşor.
relative la scalele v1 (introversiune / extraversiune) şi v2 (acceptarea normelor) din CPI.

Mulţimea S, a perechilor de scoruri obţinute la scalele v1 şi v2 de lotul experimental, utilizată pentru construirea funcţiilor F_c, este inclusă în mulţimea [12,30]x[21,33]. Această mulţime a jucat, în exemplul de faţă, rolul mulţimii XxY.

Valorile statistice ale scorurilor celor două scale sunt:

- pentru v1: m_1=21.17 σ_1=5.92
- pentru v2: m_2=26.84 σ_2=4.73

Graficele funcţiilor caracteristice F_c pe mulţimea [12,30]x[21,33] sunt prezentate în figurile 13.5.5.3 (pentru propoziţia *p2*) şi 13.5.5.5 (pentru propoziţia *p3*).

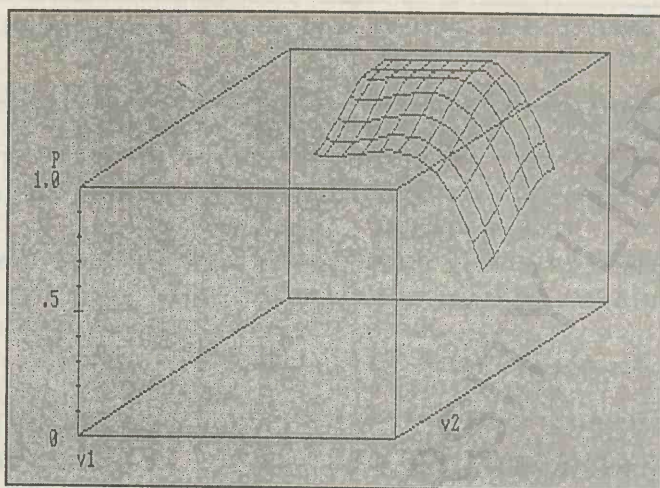


Figura 13.5.5.3
Graficul funcţiei caracteristice a propoziţiei *p2* relativă la perechea de scale (v1,v2)

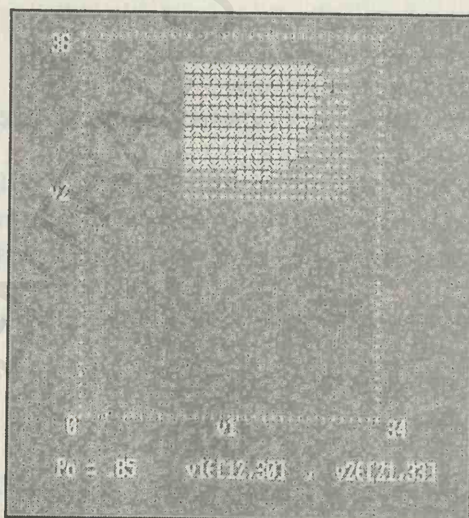


Figura 13.5.5.4
Perechile de scoruri (s_1,s_2) pentru care propoziţia *p2* se poate utiliza în descrierea persoanei

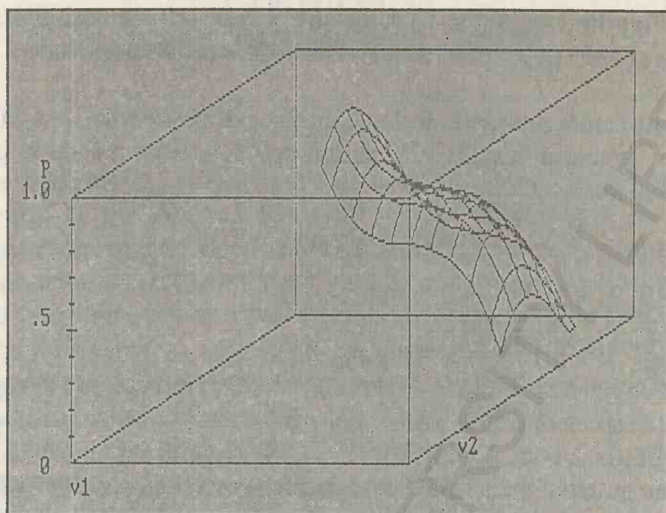


Figura 13.5.5.5
Graficul funcţiei caracteristice a propoziţiei *p3* relativă la perechea de scale (v1,v2)

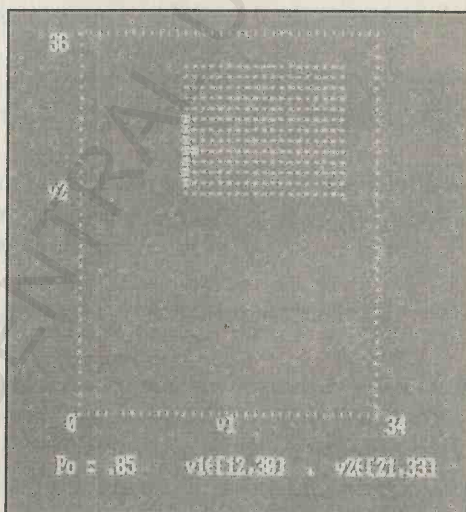


Figura 13.5.5.6
Perechile de scoruri (s_1,s_2) pentru care propoziţia *p3* se poate utiliza în descrierea persoanei

În figurile 13.5.5.4 şi 13.5.5.6 sunt notate prin cercuri pline (discuri) perechile de scoruri (s_1,s_2) pentru care valorile funcţiei caracteristice F_c sunt mai mari decât P_0.

Considerăm mulţimile formate din perechi de scoruri (s_1,s_2) (s_1=scorul la scala v1, s_2=scorul la scala v2), cărora, conform celor arătate în paragraful 13.5.1.1, li se ataşează listele 1-4:

A1 = $\{(s_1,s_2) : s_1 < m_1-\sigma_1$ şi $s_2 \geq m_2+\sigma_2\}$
A2 = $\{(s_1,s_2) : s_1 \geq m_1+\sigma_1$ şi $s_2 \geq m_2+\sigma_2\}$
A3 = $\{(s_1,s_2) : s_1 < m_1-\sigma_1$ şi $s_2 < m_2-\sigma_2\}$
A4 = $\{(s_1,s_2) : s_1 \geq m_1+\sigma_1$ şi $s_2 < m_2-\sigma_2\}$

Se observă că, pentru nici una dintre propoziţiile *p2* şi *p3*, suprafaţa pe care funcţia caracteristică ia valori mai mari decât P_0 nu coincide cu vreuna dintre mulţimile **A1∩S, A2∩S, A3∩S, A4∩S**. Aceasta înseamnă că cele două propoziţii nu ar trebui incluse în listele ataşate perechii de scale (v1,v2). Dar, ele sunt utile pentru caracterizarea persoanelor care au obţinut la cele două scale scoruri din mulţimile de puncte notate prin cercuri pline în figurile 13.5.5.4 şi 13.5.5.6.

CAPITOLUL 14
PROFILUL PSIHOLOGIC

Frecvent, atât în cercetările ştiinţifice cât şi în practică, se înregistrează valorile pe care le iau două sau mai multe variabile dependente D_1, D_2,..., D_n ($n \geq 2$) pentru diverse valori ale variabilelor independente I_1, I_2,..., I_m ($m \geq 1$). Se ajunge la următoarea problemă: dacă valorile variabilelor independente sunt alese astfel încât între oricare două înregistrări cel puţin o variabilă independentă îşi schimbă valoarea, ce se poate spune despre felul în care se modifică, de la o înregistrare la alta, ansamblul valorilor variabilelor dependente?

În acest capitol sunt discutate câteva posibilităţi de rezolvare a acestei probleme atunci când variabilele dependente D_1, D_2,..., D_n îndeplinesc următoarele condiţii:

a. sunt măsurate pe scală de acelaşi tip (de exemplu, pe scală de interval);

b. iau valori în aceeaşi mulţime (de exemplu, pot lua doar valorile 1, 2, 3, 4 şi 5);

c. fiecare valoare posibilă are aceeaşi semnificaţie pentru toate variabilele (de exemplu, dacă variabilele dependente indică prezenţa unor atribute, atunci valoarea 0 va însemna "atribut absent", iar valoarea 1, "atribut prezent").

În asemenea situaţii se obişnuieşte să se vizualizeze valorile variabilelor dependente construind câte un **profil** pentru fiecare înregistrare (Tabachnick, Fidell, 1989).

Trasarea unui profil constă în reprezentarea grafică a punctelor corespunzătoare valorilor variabilelor dependente şi unirea punctelor consecutive prin segmente de dreaptă.

Exemple de situaţii în care se pot realiza profiluri

A. Într-o cercetare referitoare la folosirea calculatoarelor într-o companie, în cursul unei luni, variabilele independente sunt compartimentul companiei şi tipul calculatorului, iar variabilele dependente reprezintă timpul mediu de utilizare şi timpul mediu de depanare al calculatoarelor. Se poate construi câte un profil pentru fiecare pereche (compartiment, tip de calculator).

B. Un director de instituţie îşi apreciază angajaţii în privinţa pregătirii profesionale, a conştiinciozităţii, a punctualităţii şi a disciplinei. Calificativele acordate

sunt numere întregi, cuprinse între 1 şi 5, cu următoarele semnificaţii: 1=foarte slab, 2=slab, 3=mediocru, 4=bun, 5=foarte bun. În acest caz, variabila independentă corespunde persoanelor din instituţie, iar variabilele dependente sunt cele patru criterii utilizate la formularea caracterizărilor. Se pot trasa atâtea profiluri, câte persoane au fost apreciate.

C. Pentru a studia eficacitatea unui medicament se fac măsurători ale unui anumit parametru fiziologic, în cinci momente (înaintea începerii tratamentului şi apoi, de patru ori, din 10 în 10 zile), utilizându-se persoane aflate în diferite stadii ale bolii. Variabila independentă are ca valori stadiile bolii, iar variabilele dependente corespund celor cinci momente de măsurare. Fiecare profil va corespunde unui stadiu al bolii. El va conţine mediile valorilor parametrului fiziologic, în cele cinci momente, calculate pe baza valorilor notate pentru toate persoanele aflate în stadiul de boală respectiv.

În practică se întâlnesc mai frecvent două moduri de trasare a profilurilor (Piéron, Pichot ş.a., 1952):

- Se foloseşte un sistem de axe ortogonale, pe una fiind notate variabilele dependente, iar pe cealaltă valorile posibile ale acestora.
- Se desenează câte o rază pentru fiecare variabilă dependentă. Razele au aceeaşi lungime, pornesc dintr-un centru comun şi oricare două raze alăturate formează între ele un unghi egal cu 360° /numărul de variabile dependente. Pe fiecare rază se marchează printr-un punct valoarea variabilei dependente.

Ordinea în care sunt notate într-un profil variabilele dependente este determinată de semnificaţia acestora. Astfel, dacă variabilele corespund unei caracteristici măsurate sau observate în mai multe momente, ordinea lor va fi cea cronologică. Dacă variabilele dependente pot fi grupate (de exemplu, unele se referă la forma şi mărimea unui produs, altele la costurile de producţie ale acestuia), atunci cele care fac parte din aceeaşi grupă vor fi reprezentate alături.

Noţiunea de profil apare şi în psihologie, mai ales în orientarea şcolară şi profesională. Variabilele independente reprezintă persoane sau grupe de subiecţi constituite după diverse criterii (de exemplu, sex, vârstă, profesie, nivel de şcolarizare etc.). Variabilele dependente sunt rezultatele unor măsurători (de exemplu, timpi de reacţie ori frecvenţele relative ale unor tipuri de erori), scorurile unor scale sau diferenţele de scor între două administrări ale unui test.

Pentru a putea construi profilul psihologic pe baza rezultatelor unor teste se obişnuieşte să se transforme cotele brute obţinute de subiecţi la diverse scale în cote standardizate, în decile sau stanine.

Atunci când profilul corespunde unui grup de persoane, valoarea fiecărei variabile dependente este media valorilor variabilei respective pentru persoanele din grup.

Trasarea profilurilor psihologice serveşte în principal următoarelor scopuri:

A. compararea între ele a unor grupe de persoane;

B. compararea unei persoane cu mai multe grupe, pentru a determina în care grupă ar trebui repartizată;

C. compararea a două persoane, una cu alta;

D. compararea unei persoane cu ea însăşi, înainte şi după un anumit eveniment (de exemplu, un tratament).

În continuare, vom prezenta, pe scurt, procedele de lucru pentru fiecare dintre aceste scopuri.

A. Compararea profilurilor grupelor utilizează metode statistice specifice, care fac obiectul "analizei de profil". Se încearcă să se dea răspuns următoarelor întrebări:

a. Cum se asociază variabilele dependente pentru fiecare grup? Ce deosebiri există între grupe în această privinţă?

b. Există vreo variabilă dependentă ale cărei valori medii diferă semnificativ între grupe?

c. Profilurile grupelor sunt paralele între ele?

Consecinţa paralelismului profilurilor este faptul că scorurile cele mai mici şi scorurile cele mai mari apar la toate grupele pentru aceleaşi variabile. Dacă grupele sunt formate din subiecţi supuşi unor tratamente diferite şi profilurile sunt paralele, se poate deduce că tratamentele au efecte similare.

d. În cazul profilurilor paralele, există diferenţe de "nivel" între grupe?

"Nivelul" unui profil este media valorilor tuturor variabilelor dependente.

B. Determinarea grupei din care face parte o persoană este o problemă de clasare, care se rezolvă adesea prin compararea persoanei respective cu prototipurile claselor şi identificarea prototipului cu care aceasta "seamănă" cel mai mult.

Atunci când profilurile grupelor sunt diferite, se poate considera fiecare profil ca fiind prototip al grupei. Se determină profilul persoanei, pe baza valorilor variabilelor dependente care au fost utilizate la trasarea profilurilor grupelor, şi se compară profilul persoanei cu profilul fiecărei grupe.

Exemplu.

Pentru a afla dacă o persoană are tulburări socio-emoţionale se pot utiliza drept variabile dependente scalele chestionarelor STAI şi STAXI. Se compară scorurile obţinute de subiect cu scorurile medii realizate de două loturi: unul format din persoane cu tulburări socio-emoţionale şi unul compus din persoane fără asemenea tulburări.

De obicei, compararea profilului persoanei cu profilul unei grupe constă în calculul distanţei dintre punctele

$$(p_1, p_2,..., p_n) \text{ şi } (g_1, g_2,..., g_n),$$

unde, pentru j=1,2,...,n:

p_j este valoarea variabilei dependente D_j la persoana respectivă;
g_j este media valorilor variabilei dependente D_j în cadrul grupei.

Pentru calculul distanţei dintre punctele $(p_1, p_2,..., p_n)$ şi $(g_1, g_2,..., g_n)$ se va utiliza una dintre distanţele spaţiului $\mathbf{R}^n$ (de exemplu, distanţa euclidiană, distanţa lui Mahalanobis, distanţa city-block etc.), aleasă în funcţie de importanţa acordată distanţelor dintre valorile componentelor (a se vedea Anexa II).

Asemănarea profilurilor este cu atât mai mare, cu cât distanţa dintre puncte este mai mică.

Compararea unui profil individual cu profilul unui grup impune să se aibă în vedere şi următoarele aspecte:

a. Rezultatele comparării profilurilor depind de funcţia aleasă pentru distanţă.

b. O aceeaşi valoare a distanţei poate să apară atunci când între profilurile comparate diferă mult valorile unei singure variabile dependente, celelalte variabile având valori egale, sau când toate variabilele dependente diferă puţin între profiluri. Sunt importante şi diferenţele mici între valorile variabilelor dependente sau acestea ar trebui neglijate? Are sau nu importanţă care este variabila dependentă pentru care apar diferenţe mari între valori?

c. În funcţiile utilizate pentru calculul distanţelor din spaţiul $\mathbf{R}^n$ nu au importanţă semnele diferenţelor p_j-g_j (j=1,2,...,n) ci doar mărimile lor în valoare absolută. Dar, în problemele practice, nu contează şi semnele diferenţelor? De exemplu, dacă se face compararea profilurilor pentru a identifica profesia cea mai potrivită unei persoane, are importanţă dacă în privinţa unei variabile individul respectiv este mai bun sau mai slab decât media grupei ori contează doar mărimea diferenţei dintre rezultatul său şi media grupei, în valoare absolută?

C şi D. Pentru a putea interpreta corect diferenţele dintre două profiluri individuale, care aparţin la doi subiecţi diferiţi sau aceluiaşi subiect în două momente diferite, ar trebui să se poată şti dacă diferenţa valorilor observate la fiecare variabilă dependentă corespunde unei deosebiri reale sau este datorată întâmplării.

Prezentăm o metodă de comparare a două profiluri individuale, care porneşte de la ideea că orice scor observat la un test este afectat de erori de măsură şi ia în considerare, pentru fiecare variabilă dependentă, scorurile care ar fi putut să apară, cu o anumită probabilitate fixată, în locul celui observat. Această metodă este aplicabilă atunci când sunt îndeplinite următoarele condiţii:

- Variabilele dependente reprezintă scoruri brute la teste care măsoară caracteristici durabile.

• Pentru fiecare test scorurile posibile aparţin unei mulţimi finite formate din numere
întregi consecutive:

$\mathbf{X}=\{x_1,x_2,...,x_n\}$ unde $x_1 \in \mathbf{Z}$ şi $x_i=x_{i-1}+1$, $i=2,3,...,n$
(aceste mulţimi pot să difere de la un test la altul).

• Fiecare test a fost administrat de două ori aceluiaşi lot de subiecţi care, în privinţa variabilelor dependente studiate nu s-au modificat între test şi retest. Verificarea acestei condiţii se face urmărind dacă relaţia dintre scorurile testului şi cele ale retestului este liniară, semnificativă statistic la un prag p fixat (p=0.01 sau p=0.001), şi dacă diferenţa mediilor scorurilor între test şi retest este nesemnificativă statistic la pragul p.

Se procedează astfel:

(1) Se fixează o probabilitate $P \in [0,1]$.

(2) Pentru fiecare test se rezolvă problema următoare (Albu, Pitariu, 1995b): să se determine, pentru fiecare scor posibil al testului, $x \in \mathbf{X}$, un interval care să îl conţină:

$$[u(x,P), v(x,P)], \quad u(x,P) \leq x \leq v(x,P),$$

astfel încât să fie îndeplinite următoarele condiţii:

(C1) Probabilitatea ca o persoană care a obţinut la o administrare a testului x puncte să realizeze la o altă administrare (anterioară sau posterioară celei la care subiectul a fost cotat cu x puncte) y puncte, $y \in [u(x,P), v(x,P)]$, să fie cel puţin egală cu P:

$$P(y \in [u(x,P),v(x,P)] \mid x) \geq P \qquad (14.1)$$

Intervalul $[u(x,P), v(x,P)]$ conţine scorurile "cele mai apropiate probabilistic" de x.

(C2) Pentru orice scor posibil $x \in \mathbf{X}$, dacă $y \in [u(x,P),v(x,P)]$, atunci şi $x \in [u(y,P),v(y,P)]$ (dacă y este unul dintre scorurile "cele mai apropiate probabilistic" de x, atunci şi x este unul dintre scorurile "cele mai apropiate probabilistic" de y);

(C3) Extremităţile intervalelor $[u(x,P),v(x,P)]$ să fie cât mai apropiate de x, pentru toate scorurile posibile x:

$$C(P)=\max_x \max \{x-u(x,P),v(x,P)-x\} \rightarrow minim \qquad (14.2)$$

Trebuie remarcat că problema determinării intervalelor $[u(x,P), v(x,P)]$, pentru $x \in \mathbf{X}$, care să respecte condiţiile (C1) şi (C2), are întotdeauna soluţie, căci

se poate construi, pentru fiecare scor posibil x, intervalul $[u(x,P),v(x,P)]$ punând $u(x,P)=x_1$ şi $v(x,P)=x_n$.

Problema construirii intervalelor $[u(x,P),v(x,P)]$, pentru $x \in X$, care să respecte condiţiile (C1), (C2) şi (C3), atunci când testul a fost administrat de două ori unui lot de persoane şi fiecare scor posibil a fost obţinut cel puţin o dată, poate fi transpusă matematic sub forma unei probleme de programare.

Se notează:

$f_{i,j}$ = numărul de subiecţi care au obţinut scorul x_i la test şi scorul x_j la retest, $i,j=1,2,...,n$;

Faptul că fiecare scor posibil a fost obţinut cel puţin o dată (indiferent la care dintre administrări) se exprimă prin inegalitatea:

$$\sum_{j=1}^{n} (f_{i,j}+f_{j,i})-f_{i,i} > 0, \quad i=1,2,...,n \tag{14.3}$$

Introducem variabilele $a_{i,j}$, $i,j=1,2,...,n$, cu următoarea semnificaţie:

$a_{i,j}=1$ dacă scorul x_i aparţine intervalului construit pentru scorul x_j şi

$a_{i,j}=0$ în caz contrar.

Se cere să se determine numerele $a_{i,j}$, $i,j=1,2,...,n$, în următoarele condiţii:

(a) $a_{i,j} \in \{0,1\}$, $i,j=1,2,...,n$ (14.4)

(din construcţia variabilelor $a_{i,j}$)

(b) $a_{i,i}=1$, $i=1,2,....,n$ (14.5)

(fiecare scor posibil x trebuie să facă parte din intervalul "scorurilor care sunt cele mai apropiate probabilistic" de el)

(c) $a_{i,j}=a_{j,i}$, $i,j=1,2,....,n$ (14.6)

(condiţia C2 a problemei)

(d) $a_{i,1} \leq a_{i,2} \leq ... \leq a_{i,i-1} \leq a_{i,i}$ (14.7)

$a_{i,i} \geq a_{i,i+1} \geq ... \geq a_{i,n-1} \geq a_{i,n}$ (14.8)

pentru $i=1,2,...,n$

(pentru a avea câte un interval $[u(x,P),v(x,P)]$ în jurul fiecărui scor posibil x)

(e)

$$P(x_i) = \frac{\sum_{j=1}^{n} a_{i,j}f_{i,j}+\sum_{j=1}^{n} a_{j,i}f_{j,i}-f_{i,i}}{\sum_{j=1}^{n} (f_{i,j}+f_{j,i})-f_{i,i}} \geq P \tag{14.9}$$

pentru i=1,2,...,n.

(condiţia C1 a problemei: $P(x_i)$ reprezintă frecvenţa relativă a subiecţilor care au obţinut la o administrare a testului scorul x_i, iar la cealaltă un scor din intervalul construit pentru x_i, din totalul celor care au obţinut la una dintre administrările testului scorul x_i)

(f)

$$C(P) = \max_{1 \le i \le n} \max \left\{ \sum_{j=i+1}^{n} a_{i,j} \,,\, \sum_{j=1}^{i-1} a_{i,j} \right\} \rightarrow \mathit{minim} \qquad (14.10)$$

(condiţia C3: prima sumă reprezintă diferenţa $v(x,P)$-x, iar a doua, diferenţa x-$u(x,P)$, pentru x=x_i)

(3) Pentru fiecare din cele două cazuri în care se trasează profilul, la fiecare test, se înlocuieşte scorul obţinut, x, cu intervalul [$u(x)$, $v(x)$], în care se află scorurile care, cu probabilitatea P, ar fi putut să apară în locul scorului x. Se înlocuieşte astfel profilul tradiţional, în care pentru fiecare test este desenat un punct, corespunzător scorului obţinut de subiect, cu un profil în care pentru fiecare test este trasat segmentul corespunzător intervalului în care, cu probabilitatea P, s-ar fi putut găsi scorul subiectului.

(4) La un test la care valorile în cele două profiluri sunt x şi y se va afirma că profilurile diferă numai dacă intervalele corespunzătoare celor două scoruri, [$u(x)$, $v(x)$] şi [$u(y)$, $v(y)$] au intersecţia vidă.

Exemplu.

În figura 14.1 sunt prezentate profilurile psihologice la CPI pentru două persoane. Fiecare profil s-a obţinut prin reprezentarea grafică a cotelor T la primele 20 de scale din CPI.

Aplicând procedeul descris mai sus, s-a determinat, pentru probabilitatea P=0.85, intervalul în care s-ar fi putut afla scorul brut al fiecărei scale. Limitele intervalelor au fost transformate în cote T. În figura 14.2 sunt prezentate profilurile celor două persoane, realizate cu ajutorul intervalelor. Se observă că pentru unele scale (Do, In, Sc, Gi, Wb, Ac) intervalele celor două profiluri au porţiuni comune, deci este posibil ca la scalele respective persoanele să nu se deosebească una de alta.

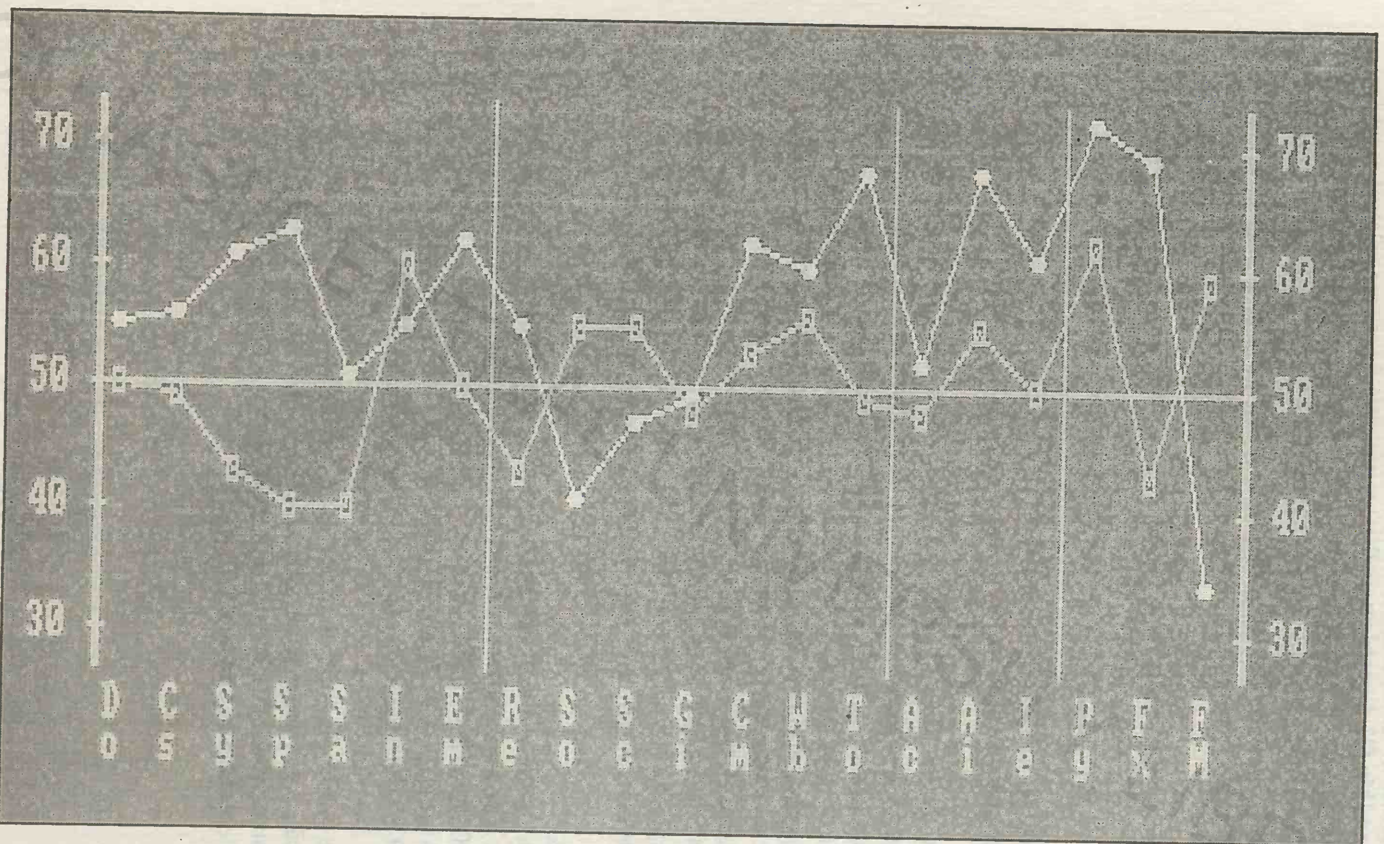


Figura 14.1
Profilurile psihologice la CPI, reprezentate cu ajutorul cotelor T, pentru două persoane

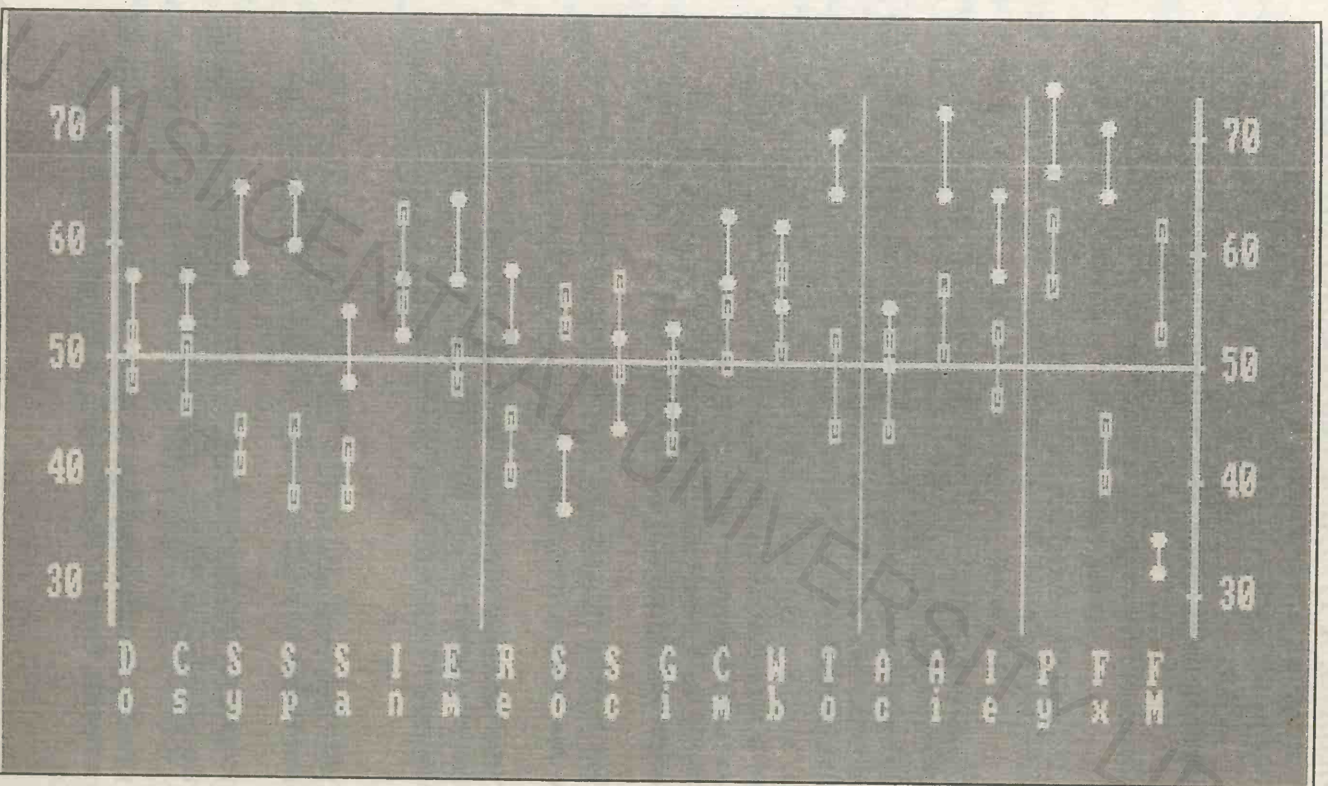


Figura 14.2
Profilurile psihologice la CPI, reprezentate cu ajutorul intervalelor corespunzătoare probabilităţii P=0.85, pentru două persoane

CAPITOLUL 15
O ABORDARE A TEORIEI TESTELOR PSIHOLOGICE BAZATĂ PE TEORIA CLASIFICĂRII

Înţelegerea acestui capitol necesită parcurgerea prealabilă a Anexei I.

15.1. TESTUL PSIHOLOGIC CA INSTRUMENT DE CLASIFICARE

Utilizarea în practică a testelor psihologice revine la rezolvarea unor probleme de clasificare. De exemplu:

A. Ordonarea persoanelor pe baza scorurilor obţinute la un test înseamnă formarea unui *clasament* în mulţimea subiecţilor.

B. Alegerea, dintr-o mulţime de candidaţi examinaţi cu mai multe probe psihologice, a celui mai potrivit pentru un anumit post impune *clasarea* candidaţilor în una din clasele disjuncte "admis" şi "respins", urmată de *ierarhizarea* celor admişi pe baza rezultatelor la probele psihologice.

C. Selecţia sau orientarea profesională se realizează conform unui algoritm compus din următorii paşi:

(1) Se stabilesc clasele în care se va face clasarea candidaţilor. O clasă corespunde unui post sau unei grupe de posturi asemănătoare prin cerinţe.

(2) Se determină variabilele, fizice sau/şi psihice, care permit o diferenţiere între posturile pentru care se face selecţia. De exemplu, pentru un post poate fi necesară rezistenţa fizică, iar pentru altul, o atenţie distributivă bună.

(3) Se aleg instrumentele care măsoară aceste variabile şi permit diferenţieri între clase, în sensul că valorile lor diferă de la o clasă la alta.

(4) Pentru fiecare clasă se descrie un *prototip*, folosind valori posibile sau ideale ale măsurătorilor făcute cu instrumentele alese.

(5) Fiecare persoană examinată cu instrumentele alese este comparată cu toate prototipurile. Ea va fi *clasată* în clasa prototipului cu care se aseamănă cel mai mult.

15.2. CONSTRUIREA TESTELOR PSIHOLOGICE

Calculul scorurilor unui test prin însumarea, ponderată sau nu, a scorurilor itemilor componenţi, reprezintă un procedeu de agregare a unor clasamente:

- Fiecare item, prin scorurile sale, defineşte o relaţie de preordine totală în mulţimea **E** a subiecţilor.
- Scorul testului, la rândul său, defineşte o relaţie de preordine totală pe aceeaşi mulţime **E** şi, deci, conduce la un clasament.

Construirea unui test psihologic urmăreşte să selecteze, dintr-o mulţime dată de itemi, un număr (fixat a priori sau nu) de itemi şi să stabilească un procedeu de agregare a clasamentelor (concret, de calcul al scorului testului) astfel încât clasamentul rezultat, corespunzător scorurilor testului, să fie cât mai asemănător cu un anumit clasament existent, cunoscut sau nu. Clasamentul existent este dat de scorurile unui criteriu sau de valorile unei variabile latente. În cel de-al doilea caz, clasamentul nu este cunoscut. Adesea se presupune că el coincide cu clasamentul realizat de scorurile factoriale ale subiecţilor pentru un factor comun itemilor.

Vom explica în continuare ideile care stau la baza procedeelor de construire a testelor.

Se consideră o mulţime de persoane, **E**, şi doi itemi de test. Pentru fiecare persoană care a răspuns la ambii itemi se calculează un "scor total", s_t, folosind formula:

$$s_t = w_1 s_1 + w_2 s_2 \qquad (15.2.1)$$

unde:

s_1, s_2 sunt scorurile persoanei la cei doi itemi;

w_1, w_2 sunt ponderi (numere reale pozitive) constante pentru toate persoanele.

Scorul total calculat astfel realizează un clasament în mulţimea **E**.

Procedeele statistice aplicate pentru construirea testelor psihologice au la bază următoarele teoreme:

TEOREMA 15.2.1. Dacă cele două clasamente corespunzătoare itemilor coincid, atunci clasamentul corespunzător scorului total coincide cu ele.

TEOREMA 15.2.2. Atunci când clasamentul corespunzător primului item este o rafinare a clasamentului corespunzător celui de-al doilea item, clasamentul corespunzător scorului total coincide cu clasamentul corespunzător primului item.

TEOREMA 15.2.3. Dacă cele două clasamente corespunzătoare itemilor coincid cu un clasament dat, atunci ele coincid între ele.

TEOREMA 15.2.4. Atunci când clasamentul corespunzător primului item coincide cu clasamentul corespunzător celui de-al doilea item, care este o rafinare a unui clasament dat, clasamentul corespunzător primului item este şi el o rafinare a clasamentului dat.

Aceste teoreme se aplică succesiv, luând în locul celor doi itemi, testul existent la un moment dat şi un item din mulţimea celor care pot fi adăugaţi testului.

În practica psihologică, însă, se întâlnesc foarte rar clasamente care să coincidă sau unul să fie o rafinare a altuia. Intervin diverse erori de măsură care au drept consecinţă faptul că persoane care obţin acelaşi scor la un item (şi, deci, aparţin aceleiaşi clase în partiţia ordonată corespunzătoare itemului) obţin scoruri diferite la un alt item (făcând parte din clase diferite în partiţia ordonată corespunzătoare acestuia). Din acest motiv nu se mai caută clasamente astfel încât unul să fie o rafinare a altuia sau să coincidă cu el, ci clasamente care să "concorde" între ele. Concret, ce se înţelege prin "concordanţă", diferă de la o metodă statistică la alta. Dar, în esenţă, concordanţa a două clasamente se referă la apropierea lor de situaţia când ele coincid (dacă relaţiile de preordine totală cărora le corespund clasamentele au acelaşi număr de clase de echivalenţă) sau când unul este o rafinare a celuilalt (dacă numărul de clase de echivalenţă ale relaţiilor de preordine totală diferă).

În cadrul analizei de itemi, al metodelor de construire a testelor şi al cercetării validităţii relative la criteriu se folosesc diverse formule pentru evaluarea concordanţei a două clasamente.

Astfel, studiul validităţii relative la criteriu a testelor şi al validităţii itemilor compară clasamentul corespunzător unui criteriu cu cel corespunzător testului sau itemului cu ajutorul unui coeficient de corelaţie (corelaţie liniară, corelaţie biserială sau corelaţia rangurilor).

În cercetarea puterii de discriminare a itemilor, clasificarea corespunzătoare criteriului se reduce la trei clase, din care, pentru verificarea concordanţei clasamentelor se folosesc doar cele două extreme ("contrastante"). Itemul are putere de discriminare dacă scorurile sale realizează un clasament concordant cu cel dat de clasele contrastante.

Eliminarea itemilor al căror indice de dificultate este egal sau aproximativ egal cu 0 sau cu 1 are ca explicaţie faptul că un clasament în care toate persoanele ocupă acelaşi loc nu poate să concorde cu nici un clasament în care există mai multe locuri.

Construirea unui test care să aibă o validitate bună relativ la un criteriu, prin selectarea acelor itemi care au o validitate bună faţă de criteriul respectiv, este de asemenea o aplicaţie a teoremelor 15.2.1 şi 15.2.3, în care termenul "coincid" a fost înlocuit cu termenul "concordă". Astfel: atunci când clasamentele corespunzătoare celor doi itemi concordă cu clasamentul corespunzător criteriului, clasamentele lor concordă între ele, va concorda şi cu clasamentul corespunzător scorului total care, la rândul său, va concorda cu clasamentul corespunzător criteriului.

Importanţa introducerii în test doar a itemilor valizi rezultă din următorul exemplu.

Exemplu.

Să considerăm 10 persoane, $s_1, s_2,...,s_{10}$, pentru care se cunosc valorile unui criteriu. Să presupunem că scorurile criteriului verifică inegalităţile:

$$c(s_1) < c(s_2) < ... < c(s_{10}) \qquad (15.2.2)$$

Se doreşte să se construiască un test care să fie valid faţă de acest criteriu.

Se porneşte de la o colecţie iniţială formată din 6 itemi dihotomici, $i_1, i_2,...,i_6$. Scorurile obţinute de subiecţi la aceşti itemi şi la testul compus din ei sunt notate în Tabelul 15.2.1.

Tabelul 15.2.1

Scorurile itemilor $i_1, i_2,...,i_6$ pentru subiecţii $s_1, s_2,...,s_{10}$

Item	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
i_1	0	0	0	0	0	0	0	0	1	1
i_2	0	0	0	0	0	0	0	1	1	1
i_3	0	0	0	1	1	1	1	1	1	1
i_4	0	0	1	1	1	1	1	0	0	0
i_5	1	1	1	1	1	1	0	0	0	0
i_6	1	1	1	1	0	0	0	0	0	0
Scor total	2	2	3	4	3	3	2	2	3	3

Analizând scorurile itemilor şi ale testului se deduc următoarele:

• Pentru primii trei itemi partiţiile ordonate corespunzatoare scorurilor lor sunt rafinate de partiţia ordonată corespunzatoare scorurilor criteriului. Dacă un subiect s_i este inferior altui subiect s_j la criteriu, $c(s_i)<c(s_j)$, atunci s_i nu obţine la item un scor mai mare decât s_j ($i,j \in \{1,2,...,10\}$).

• Itemii i_5 şi i_6 produc clasamente "inverse": dacă s-ar schimba modul lor de cotare (s-ar da 1 punct în loc de 0 puncte şi invers), ar rezulta clasamente pentru care clasamentul bazat pe valorile variabilei criteriu ar fi o rafinare.

• Itemul i_4 este necorespunzător: obţin scorul 1 la item persoanele care au rezultate medii la criteriu şi, prin urmare, clasamentul produs de el nu coincide şi nu este rafinat de clasamentul produs de criteriu.

• Testul format din cei şase itemi produce o ierarhie foarte diferită de cea bazată pe variabila criteriu. De exemplu, scorul maxim la test îl are subiectul s_4 care nu se află printre cei cu valorile cele mai mari ale criteriului.

Dar, ţinând cont de rezultatele analizei validităţii itemilor, se poate construi un test care să nu modifice prea mult ordinea subiecţilor. Pentru aceasta se schimbă modul de cotare pentru itemii i_5 şi i_6 (scor nou := 1- scor vechi) şi se renunţă la itemul i_4. Scorurile noi ale itemilor şi scorul total al testului sunt prezentate în tabelul 15.2.2.

Tabelul 15.2.2

Scorurile noi ale itemilor şi scorul testului pentru subiecţii $s_1, s_2,...,s_{10}$

Item	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
i_1	0	0	0	0	0	0	0	0	1	1
i_2	0	0	0	0	0	0	0	1	1	1
i_3	0	0	0	1	1	1	1	1	1	1
i_5	0	0	0	0	0	0	1	1	1	1
i_6	0	0	0	0	1	1	1	1	1	1
Scor total	0	0	0	1	2	2	3	4	5	5

În teoria răspunsului la itemi se consideră că un item dihotomic este valid dacă are curba caracteristică (curba de regresie) crescătoare faţă de variabila măsurată de test. Alura curbei de regresie constituie, deci, o metodă de verificare a concordanţei între clasamentul corespunzător scorurilor itemului şi cel corespunzător valorilor variabilei măsurate de itemi (acest al doilea clasament nu este cunoscut).

Şi atunci când testul se construieşte utilizând teoria răspunsului la itemi, obţinerea unui clasament corect al subiecţilor este condiţionată de alegerea itemilor valizi.

Exemplu (Albu, 1993).

Considerăm un test alcătuit din trei itemi dihotomici, care măsoară aceeaşi trăsătură psihică şi ale căror funcţii caracteristice au expresia analitică:

$$P(y) = c + \frac{1-c}{1+e^{-1,7a(y-b)}} \qquad (15.2.3)$$

Parametrii funcţiilor caracteristice sunt :

$a_1=1$ $b_1=6$ $c_1=0$
$a_2=-0.1$ $b_2=15$ $c_2=0$
$a_3=1$ $b_3=11$ $c_3=0$

Funcţia caracteristică a celui de-al doilea item este descrescătoare ($a_2<0$), deci itemul nu este valid.

La fiecare nivel al variabilei măsurate, $t\in\mathbf{R}$, sunt posibile, teoretic, toate scorurile testului. Luând în considerare probabilităţile scorurilor, există în fiecare punct $t\in\mathbf{R}$ o repartiţie de probabilitate de următoarea formă:

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ P_0(t) & P_1(t) & P_2(t) & P_3(t) \end{pmatrix} \qquad (15.2.4)$$

cu $P_i(t) \geq 0$, $i=0,1,2,3$
şi $P_0(t)+P_1(t)+P_2(t)+P_3(t)=1$.

Valorile $P_0(t)$, $P_1(t)$, $P_2(t)$ şi $P_3(t)$ se pot calcula cu ajutorul funcţiilor caracteristice ale itemilor.

Notăm cu $A_i(t)$ evenimentul ca itemul i să fie cotat cu un punct pentru o persoană având nivelul variabilei măsurate egal cu t, pentru $i=1,2,3$. Probabilitatea evenimentului $A_i(t)$ este egală cu valoarea în punctul t a funcţiei caracteristice a itemului, $p_i(t)$.

Acceptăm ipotezele obişnuite din teoria răspunsului la itemi (Lord, 1980):

- variabila măsurată ia valori pe întreaga axă reală;
- în fiecare punct $t\in\mathbf{R}$, evenimentele $A_i(t)$, $i=1,2,3$, sunt independente probabilistic între ele.

În acest caz se pot scrie formule de calcul pentru probabilităţile P_i, $i=0,1,2,3$. Acestea sunt (pentru simplificarea scrierii renunţăm la notarea argumentului t):

$P_0=(1-p_1)(1-p_2)(1-p_3)$
$P_1=p_1(1-p_2)(1-p_3)+(1-p_1)p_2(1-p_3)+(1-p_1)(1-p_2)p_3$
$P_2=(1-p_1)p_2p_3+p_1(1-p_2)p_3+p_1p_2(1-p_3)$
$P_3=p_1p_2p_3$

S-a studiat testul format din cei trei itemi cu ajutorul funcţiei care ia ca valoare, în fiecare punct $t\in\mathbf{R}$, scorul cel mai probabil la test, $S_{max}(t)$:

$$S_{max}:\mathbf{R} \to \{0,1,2,3\}$$

$S_{max}(t)= \max\{j \mid P_j(t)=\max\{P_i(t) \mid i=0,1,2,3\}\}$

($S_{max}(t)$ este scorul de test care are cea mai mare probabilitate în punctul t; dacă există mai multe valori pentru care probabilitatea este egală cu cea maximă, atunci $S_{max}(t)$ este cea mai mare dintre ele.)

Graficul funcţiei S_{max} este prezentat in figura 15.2.1. După cum se observă, scorul cel mai probabil al testului nu mai este nedescrescător faţă de caracteristica măsurată de test. Persoanele pentru care scorul cel mai probabil este scorul maxim (S_{max}=3) sunt cele cu un nivel mediu al variabilei măsurate de test.

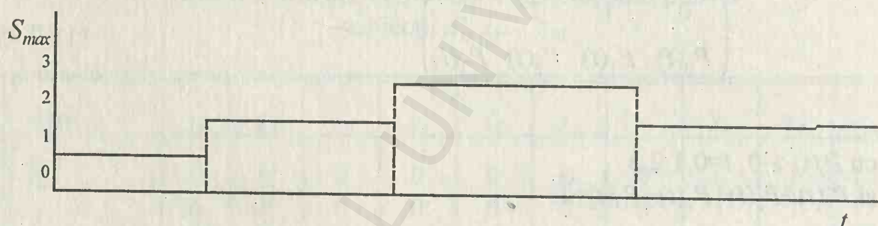


Figura 15.2.1
Graficul funcţiei S_{max}

Din exemplele prezentate se deduce că, pentru a obţine un test care să realizeze un clasament al subiecţilor concordant cu cel dat de un criteriu sau de o variabilă latentă este necesar să se analizeze validitatea itemilor şi să se reţină în test doar itemi valizi.

Metoda analizei factoriale construieşte un test psihologic selectând itemii cu saturaţii mari într-un acelaşi factor, adică itemii care corelează liniar semnificativ cu un acelaşi factor. O altă metodă de construire a testelor reţine în test doar itemii care corelează liniar semnificativ unii cu alţii. Ideea care stă la baza acestor procedee este că există o variabilă latentă ale cărei valori produc un clasament cu care trebuie să concorde clasamentul corespunzător scorurilor testului. În primul caz, se face presupunerea că factorul comun itemilor produce un clasament concordant celui dat de variabila latentă. În cel de-al doilea caz, nu se ştie nimic despre clasamentul produs de variabila latentă dar se presupune

că toţi itemii ale căror clasamente sunt concordante între ele produc clasamente concordante cu cel al variabilei latente.

15.3. VALIDITATEA UNUI TEST

15.3.1. Validitatea relativă la construct

Aprecierea validităţii unui test se bazează pe corectitudinea interpretărilor date scorurilor testului şi a concluziilor desprinse din acestea. În majoritatea cazurilor, pentru interpretarea rezultatelor unei testări se compară scorurile fiecărei persoane examinate cu scorurile altor persoane. Comparaţia înseamnă să se identifice care dintre relaţiile >, = sau < este prezentă între două scoruri.

Pentru a se putea face inferenţe corecte despre persoane, plecând de la rezultatele unui test, este necesar ca relaţiile >, = sau < existente între scorurile testului să reflecte corect relaţia existentă între valorile variabilei măsurate de test. S-a folosit aici expresia "să reflecte corect" în locul expresiei "să fie identică" deoarece, după cum se va arăta în continuare, identitatea relaţiilor nu este, în general, posibilă. Relaţiile dintre scorurile testului "reflectă corect" relaţiile dintre valorile variabilei măsurate de test dacă ierarhia subiecţilor constituită pe baza scorurilor testului aproximează bine (într-un sens care va fi precizat în continuare) ierarhia existentă între aceste persoane în privinţa variabilei respective.

Notăm:

A = mulţimea persoanelor examinate cu un test;

X = mulţimea scorurilor posibile ale testului:

$$\mathbf{X} = \{x_1, x_2, \ldots, x_n\}, \text{ cu } x_1 < x_2 < \ldots < x_n$$

$t(a)$ = scorul la test pentru persoana $a \in \mathrm{A}$;

$t(\mathrm{A}) = \{t(a) | a \in \mathrm{A}\}$

Presupunem că mulţimile **A** şi **X** sunt finite.

Definim în mulţimea A trei relaţii, astfel:

$a >_T b \iff t(a) > t(b),\ a,b \in \mathrm{A}$

$a =_T b \iff t(a) = t(b),\ a,b \in \mathrm{A}$

$a \geq_T b \iff a >_T b$ sau $a =_T b,\ a,b \in \mathrm{A}$

Relaţia $=_T$ este o relaţie de echivalenţă, iar relaţia $\geq_T$ este o relaţie de preordine totală. Prin relaţia $\geq_T$, testul realizează un clasament al persoanelor examinate.

În teoria răspunsului la itemi se consideră că variabila psihică măsurată de test poate lua orice valoare reală între $-\infty$ şi $+\infty$.

Notăm :

$\Theta(a)$ = valoarea variabilei psihice măsurate de test, pentru persoana $a \in A$;

$\Theta(A)=\{\Theta(a)|a\in A\}$

Întrucât valorile $\Theta(a)$, $a\in A$, sunt numere reale, elementele mulţimii $\Theta(A)$ pot fi comparate cu ajutorul relaţiilor >, ≥ şi = existente în R. În mulţimea A vom defini trei relaţii binare astfel:

$a>_S b \iff \Theta(a)>\Theta(b),\ \ a,b\in A$

$a=_S b \iff \Theta(a)=\Theta(b),\ \ a,b\in A$

$a\geq_S b \iff a>_S b$ sau $a=_S b,\ \ a,b\in A$

Relaţia $=_S$ este o relaţie de echivalenţă, iar relaţia $\geq_S$ este o relaţie de preordine totală.

Deci, în mulţimea A avem două clasamente, unul produs de relaţia $\geq_T$ şi unul produs de relaţia $\geq_S$.

Întrucât variabila psihică poate lua orice valoare reală, în timp ce mulţimea scorurilor posibile ale testului, X, este finită, în mod obişnuit numărul claselor de echivalenţă ale relaţiei de preordine totală $\geq_S$ este mai mare decât cel al claselor de echivalenţă în relaţia de preordine totală $\geq_T$.

Considerăm că un test care are funcţia de măsurare a unei variabile psihice este un instrument de măsură valid dacă îndeplineşte următoarele două condiţii:

a. oricare două persoane care au aceeaşi valoare a variabilei psihice măsurate obţin la test acelaşi rezultat:

$$a=_S b \Rightarrow a=_T b$$

b. orice persoană poate fi superioară alteia la test numai dacă îi este superioară şi în privinţa variabilei criteriu:

$$a>_T b \Rightarrow a>_S b$$

Aceste condiţii conduc la următoarea formulare:

DEFINIŢIA 15.3.1.1. Un test este **valid** pentru mulţimea de persoane A atunci când clasamentele corespunzătoare relaţiilor de preordine totală $\geq_S$ şi $\geq_T$ coincid sau clasamentul corespunzător relaţiei $\geq_S$ este o rafinare a clasamentului corespunzător relaţiei $\geq_T$.

Înseamnă că în cazul unui test valid în sensul acestei definiţii oricare două persoane care au aceeaşi valoare pentru variabila psihică măsurată vor avea rezultate egale la test. Deci se poate construi o funcţie

$$F: R \to X$$

$F(\Theta)$ = rezultatul obţinut la test de toate persoanele care au valoarea Θ pentru variabila psihică,

care are următoarea proprietate:

$$\Theta_1 > \Theta_2 \Rightarrow F(\Theta_1) \geq F(\Theta_2))$$

În realitate, însă, cu excepţia testelor care au un singur rezultat posibil, ceea ce face să nu prezinte interes practic, nu se poate construi niciodată o asemenea funcţie. Deci nici un test pentru care numărul scorurilor posibile este mai mare decât 1 nu satisface condiţiile definiţiei 15.3.1.1. Aceasta, deoarece nici un test nu realizează o măsurare perfectă: din cauza a numeroşi factori care intervin (condiţiile de administrare şi de cotare, starea de oboseală sau boală a celor examinaţi, hazardul etc.) pentru o valoare a variabilei psihice măsurate Θ se pot obţine mai multe rezultate de test. Mai exact, pentru fiecare valoare reală Θ există o repartiţie de probabilitate a scorurilor testului:

$$\begin{pmatrix} x_1 & x_2 & \dots & x_n \\ P_1(\Theta) & P_2(\Theta) & \dots & P_n(\Theta) \end{pmatrix} \qquad (15.3.1.1)$$

cu $0 \leq P_i(\Theta) \leq 1$, $i=1,2,\dots,n$
şi $P_1(\Theta)+P_2(\Theta)+\dots+P_n(\Theta)=1$,
unde $P_i(\Theta)$ reprezintă probabilitatea scorului x_i, $i=1,2,\dots,n$.

Vom încerca să redefinim validitatea testului utilizînd funcţia care are ca valoare, în fiecare punct $\Theta \in \mathbf{R}$, scorul cel mai probabil la test, $S_{max}(\Theta)$.

$$S_{max} : \mathbf{R} \rightarrow \mathbf{X}$$

$$S_{max}(\Theta) = \max \{x_j \in \mathbf{X} | P_j(\Theta) = \max\{P_i(\Theta) | i=1,2,\dots,n\}\} \qquad (15.3.1.2)$$

($S_{max}(\Theta)$ este valoarea de test care are cea mai mare probabilitate în punctul Θ; dacă există mai multe valori pentru care probabilitatea este egală cu cea maximă, atunci $S_{max}(\Theta)$ este cea mai mare dintre ele.)

Considerăm că un test este un instrument de măsură bun numai dacă îndeplineşte următoarele condiţii:

(C1) două persoane care au aceeaşi valoare a variabilei psihice au şanse mari să obţină acelaşi rezultat la test;

(C2) există şanse mari ca dintre două persoane diferite în privinţa variabilei psihice, cea care are o valoare mai mare a acestei variabile să realizeze la test un rezultat cel puţin la fel de bun ca şi cealaltă persoană.

Se observă că, dacă sunt îndeplinite condiţiile:

- pentru fiecare valoare $\theta \in \mathbf{R}$:
  - mulţimea $\{x_j \in \mathbf{X} | P_j(\Theta) = \max\{P_i(\Theta) | i=1,2,\dots,n\}\}$ conţine un singur element şi
  - probabilitatea rezultatului x_j, $P_j(\theta)$, este mult mai mare decât probabilităţile celorlalte scoruri de test în punctul θ;

- funcţia S_{max} este nedescrescătoare:

$\theta_1 > \theta_2 \Rightarrow S_{max}(\theta_1) \geq S_{max}(\theta_2)$ (15.3.1.3)

atunci sunt verificate condiţiile (C1) şi (C2) de mai sus.

Atunci când în fiecare punct $\theta \in R$ există un singur scor cu probabilitate mare, iar această probabilitate este mult mai mare decât probabilităţile celorlalte scoruri, funcţia S_{max} este apropiată de funcţia S_{med}, care ia ca valoare, în fiecare punct $\theta \in R$ scorul mediu al testului:

$$S_{med}: R \to [x_1, x_n]$$

$$S_{med}(t) = \sum_{i=1}^{n} x_i P_i(t) \qquad (15.3.1.4)$$

Pentru măsurarea distanţei dintre funcţiile S_{max} şi S_{med} se poate folosi norma spaţiului L_1:

$$\|S_{max} - S_{med}\| = \int_{-\infty}^{+\infty} |S_{max}(t) - S_{med}(t)|\, dt \qquad (15.3.1.5)$$

Cu ajutorul funcţiei S_{max} putem defini trei relaţii binare în mulţimea A:

$a >_P b \Leftrightarrow S_{max}(\theta(a)) > S_{max}(\theta(b)), \quad a,b \in A$

$a =_P b \Leftrightarrow S_{max}(\theta(a)) = S_{max}(\theta(b)), \quad a,b \in A$

$a \geq_P b \Leftrightarrow a >_P b$ sau $a =_P b, \quad a,b \in A$

Relaţia $=_P$ este o relaţie de echivalenţă, iar relaţia $\geq_P$ este o relaţie de preordine totală.

Dacă este îndeplinită condiţia (15.3.1.3) atunci au loc implicaţiile:

$a =_S b \Leftrightarrow \theta(a) = \theta(b) \Rightarrow a =_P b$

$a >_P b \Leftrightarrow S_{max}(\theta(a)) > S_{max}(\theta(b)) \Rightarrow \theta(a) > \theta(b) \Leftrightarrow a >_S b$

Se deduce că, dacă este îndeplinită condiţia (15.3.1.3) atunci clasamentul corespunzător relaţiei $\geq_S$ coincide cu clasamentul corespunzător relaţiei $\geq_P$ sau este o rafinare a acestuia.

DEFINIŢIA 15.3.1.2. Vom spune că un test este **P-valid** pentru mulţimea persoanelor examinate A atunci când clasamentele corespunzătoare relaţiilor de preordine totală $\geq_S$ şi $\geq_P$ coincid sau clasamentul corespunzător relaţiei $\geq_S$ este o rafinare a clasamentului corespunzător relaţiei $\geq_P$.

Condiţia de P-validitate este mai slabă decât condiţia de validitate formulată prin definiţia 15.3.1.1. Dacă un test este valid, atunci pentru orice valoare $\theta \in R$ există un singur scor de test cu probabilitatea diferită de zero, care

corespunde valorii funcţiei **F** în punctul θ, deci funcţia S_{max} coincide cu funcţia F şi este verificată condiţia (15.3.1.3), adică testul este P-valid. Dar nu orice test P-valid este valid.

Se pune, în mod firesc, întrebarea: cum se poate construi un test P-valid? Având o mulţime de itemi **M**, ale căror funcţii caracteristice p_i, i=1,2,...,n, sunt cunoscute, care itemi ar trebui aleşi din **M** pentru a rezulta un test P-valid?

Se fac următoarele convenţii, obişnuite în teoria răspunsului la itemi:

a. toţi itemii testului măsoară aceeaşi variabilă psihică;

b. la orice valoare $\theta \in$ R fixată, modul de rezolvare a unui item este independent de modul de rezolvare a celorlalţi itemi ai testului.

Au fost determinate următoarele condiţii referitoare la funcţiile caracteristice ale itemilor, care asigură obţinerea unui test P-valid.

TEOREMA 15.3.1.1. (Albu, Pitariu, 1991a). Dacă funcţiile p_i, i=1,...,n sunt derivabile şi nedescrescătoare pe **R**, $p_i(-\infty)=0$, $p_i(+\infty)=1$ pentru i=1,2,...,n atunci există k+1 puncte ($0 < k < n$-1):

$$-\infty < t_0 < t_1 < \ldots < t_k < +\infty$$

astfel încât:

$$S_{\max}(\theta) = \begin{cases} 0, & \theta < t_0 \\ i_1, & \theta \in [t_0, t_1) \\ i_2, & \theta \in [t_1, t_2) \\ \ldots & \\ i_k, & \theta \in [t_{k-1}, t_k) \\ n, & \theta \geq t_k \end{cases} \qquad (15.3.1.6)$$

cu $0 < i_1 < i_2 < \ldots < i_k < n$.

Deci, în aceste condiţii, funcţia S_{max} este o funcţie în scară, nedescrescătoare. Lungimile "treptelor" scării (adică a segmentelor care corespund câte unei valori de scor) sunt determinate tot de funcţiile p_i. Prezentăm, în acest sens, următoarea teoremă.

TEOREMA 15.3.1.2. (Albu, Pitariu, 1992a). Dacă funcţiile p_i, i=1,2,...,n, sunt continue pe **R** şi dacă există punctele

$$T_{1,min} < T_{1,max} \leq T_{2,min} < T_{2,max} \leq \ldots \leq T_{n,min} < T_{n,max}$$

astfel încât

$$p_i(\theta)\begin{cases} =0, & \theta\in(-\infty,T_{i,\min}] \\ \in(0,1), & \theta\in(T_{i,\min},T_{i,\max}) \\ =1, & \theta\in[T_{i,\max},+\infty) \end{cases}$$

pentru i=1,2,...,n, fiecare funcţie p_i fiind crescătoare pe intervalul $(T_{i,min}, T_{i,max})$, atunci există n puncte $t_i \in (T_{i,min}, T_{i,max})$, i=1,2,...,n, cu proprietatea că

$$S_{\max}(\theta) = \begin{cases} 0, & \theta<t_1 \\ k, & \theta\in[t_k,t_{k+1}),\ k=1,2,...,n-1 \\ n, & \theta\geq t_n \end{cases}$$

Punctele t_i sunt soluţiile ecuaţiilor $p_i(\theta) = 0.5$ pentru i=1,2,...,n.

Condiţiile impuse de teorema 15.3.1.2 asupra funcţiilor p_i şi graficul funcţiei S_{max} corespunzătoare sunt ilustrate în figura 15.3.1.1.

Pentru un test ai cărui itemi satisfac ipotezele teoremei 15.3.1.2 are loc egalitatea:

$$\|S_{\max}-S_{med}\| = \sum_{k=1}^{n}\left\{\int_{T_{k,\min}}^{t_k} p_k(t)dt + \int_{t_k}^{T_{k,\max}} (1-p_k(t))dt\right\} \tag{15.3.1.7}$$

Această cantitate este reprezentată de suma ariilor suprafeţelor haşurate din Figura 15.3.1.1.

Pentru minimizarea valorii $\|S_{max}-S_{med}\|$ se vor alege itemi ale căror funcţii caracteristice au intervale $[T_{k,min}, T_{k,max}]$ cu lungimi mici.

Un test ai cărui itemi satisfac ipotezele teoremei 15.3.1.2 prezintă încă un avantaj important: pe fiecare interval $I_k=\{\theta\in\mathrm{R}|S_{max}(\theta)=k\}$, k=1,2,...,n-1, pe lângă scorul k, care are probabilitatea maximă, au probabilitate diferită de zero doar scorurile k-1 şi k+1.

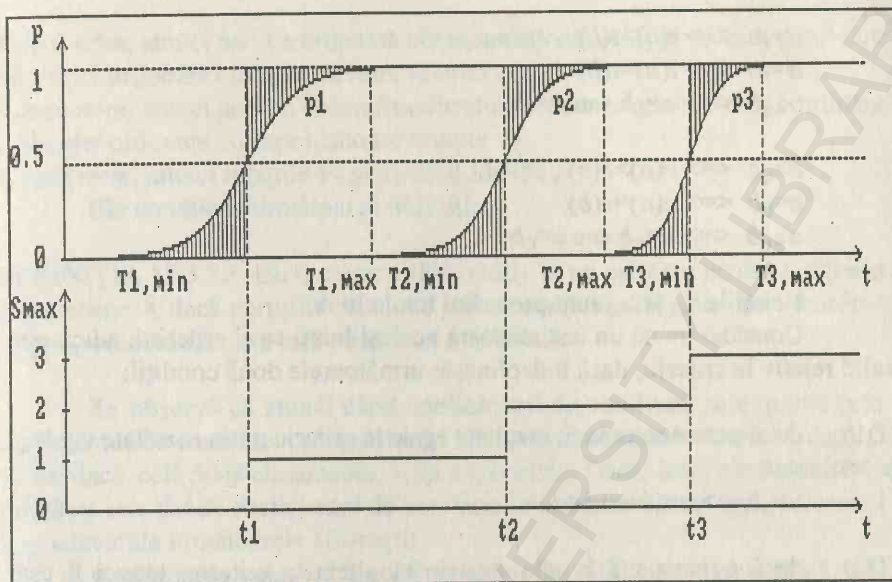


Figura 15.3.1.1
Legătura dintre funcţia S_{max} şi funcţiile caracteristice ale itemilor

15.3.2. Validitatea relativă la criteriu

Validarea unui test relativă la un criteriu constă în compararea ierarhiei produse de scorurile testului cu o altă ierarhie, realizată de o variabilă criteriu. Adică, validarea relativă la criteriu constă în compararea a două clasamente.

Notăm:

A = o mulţime formată din N persoane care au fost examinate cu testul a cărui validitate se studiază şi pentru care se cunosc valorile variabilei criteriu;

T = mulţimea scorurilor de test obţinute de persoanele din mulţimea A: **T** = $\{t_1, t_2,...,t_n\}$, $t_i \in$R pentru i=1,2,...,n şi $t_1<t_2<...<t_n$

C = mulţimea valorilor criteriului pentru persoanele din mulţimea A; **C** = $\{c_1, c_2,...,c_m\}$, $c_j \in$R pentru j=1,2,...,m şi $c_1<c_2<...<c_m$

$t(a)$ = scorul la test al unei persoane $a \in$A;

$c(a)$ = valoarea criteriului pentru persoana $a \in$A;

Definim în mulţimea A următoarele relaţii:

pentru orice $a,b \in$A:

$a >_T b \iff t(a) > t(b)$
$a =_T b \iff t(a) = t(b)$
$a \geq_T b \iff a >_T b$ sau $a =_T b$

$a >_C b \iff c(a) > c(b)$
$a =_C b \iff c(a) = c(b)$
$a \geq_C b \iff a >_C b$ sau $a =_C b$

Relaţiile $\geq_T$ şi $\geq_C$ sunt preordini totale în A.

Considerăm că un test măsoară acelaşi lucru ca şi criteriul, adică este valid relativ la criteriu, dacă îndeplineşte următoarele două condiţii:

(D1) două persoane care au rezultate egale la criteriu obţin rezultate egale şi la test:

$a =_C b \Rightarrow a =_T b \quad a,b \in A$

(D2) dacă o persoană îi este superioară alteia la criteriu, atunci îi este superioară şi la test:

$a >_C b \Rightarrow a >_T b \quad a,b \in A$

Cele două condiţii impun ca în mulţimea A clasamentele $\geq_T$ şi $\geq_C$ să fie identice:

$a \geq_T b \iff a \geq_C b, \quad a,b \in A.$

Dar acest lucru înseamnă că numărul claselor de echivalenţă în raport cu relaţiile $=_T$ şi $=_C$ este acelaşi. Fiecare clasă de echivalenţă pentru relaţia $=_T$ corespunde unei valori de test, iar fiecare clasă de echivalenţă pentru relaţia $=_C$ corespunde unei valori a criteriului. Or, în multe situaţii întîlnite în practică, numărul valorilor posibile ale testului diferă de numărul valorilor posibile ale criteriului, ceea ce face ca să nu poată fi îndeplinite condiţiile (D1) şi (D2). Din acest motiv, am înlocuit aceste condiţii cu următoarele:

- dacă $n<m$ atunci:
  - (D1.1) $a =_C b \Rightarrow a =_T b$
  - (D2.1) $a >_C b \Rightarrow a \geq_T b$
- dacă $n>m$, atunci:
  - (D1.2) $a =_T b \Rightarrow a =_C b$
  - (D2.2) $a >_T b \Rightarrow a \geq_C b$
- dacă $n=m$, atunci:
  - (D1.3) $a =_C b \Rightarrow a =_T b$
  - (D2.3) $a >_C b \Rightarrow a >_T b$

Aceste condiţii se pot transcrie astfel:

● dacă $n<m$, atunci partiţia ordonată corespunzătoare relaţiei $\geq_C$ este o rafinare a partiţiei ordonate corespunzătoare relaţiei $\geq_T$;
● dacă $n>m$, atunci partiţia ordonată corespunzătoare relaţiei $\geq_T$ este o rafinare a partiţiei ordonate corespunzătoare relaţiei $\geq_C$;
● dacă $n=m$, atunci relaţiile $\geq_T$ şi $\geq_C$ sunt identice.

Ele conduc la următoarea definiţie:

DEFINIŢIA 15.3.2.1. Un test este **valid relativ la un criteriu** pentru mulţimea de persoane A dacă partiţiile ordonate corespunzătoare relaţiilor de preordine $\geq_T$ şi $\geq_C$ coincid sau una este o rafinare a celeilalte.

Se observă că atunci când coeficientul de validitate se exprimă prin coeficientul de corelaţie a rangurilor al lui Spearman, el ia valoarea 1 dacă şi numai dacă cele două clasamente, $\geq_C$ şi $\geq_T$, coincid. Dacă, însă, coeficientul de validitate este dat de coeficientul de corelaţie liniară între test şi criteriu, atunci sunt adevărate următoarele afirmaţii:

● obţinerea valorii 1 a coeficientului de validitate înseamnă că sunt verificate condiţiile din definiţia 15.3.2.1;

● nu întotdeauna când sunt satisfăcute condiţiile definiţiei 15.3.2.1 coeficientul de validitate ia valoarea 1.

Deci coeficientul de corelaţie liniară nu ne permite să deducem dacă un test este valid relativ la criteriu în sensul definiţiei 15.3.2.1.

În continuare propunem o metodă de analiză a relaţiei test-criteriu, utilă pentru aprecierea validităţii relative la criteriu a unui test atunci când mulţimea A conţine un număr mare de persoane.

Pe baza valorilor variabilei criteriu şi a valorilor obţinute la test de către subiecţii din mulţimea A se alcătuieşte următorul tabel de frecvenţe:

Test	Criteriu $c_1 \dots c_j \dots c_m$
t_1	$f_{1,1} \cdots f_{1,j} \cdots f_{1,m}$
...	...
t_i	$f_{i,1} \cdots f_{i,j} \cdots f_{i,m}$
...	...
t_n	$f_{n,1} \cdots f_{n,j} \cdots f_{n,m}$

unde $f_{i,j}$ = numărul de subiecţi care au valoarea c_j la criteriu şi scorul t_i la test, $i=1,2,\dots,n$; $j=1,2,\dots,n$.

Din felul în care au fost definite mulţimile **C** şi **T** rezultă:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} f_{i,j} = N$$

$$\sum_{i=1}^{n} f_{i,j} > 0, \quad j=1,2,...,m$$

$$\sum_{j=1}^{m} f_{i,j} > 0, \quad i=1,2,...,n$$

Pentru fiecare pereche de valori de test $t_i, t_j \in T$, $i<j$, se calculează probabilitatea ca un subiect care a obţinut rezultatul t_i să fie inferior, în privinţa variabilei criteriu, unei persoane care a realizat la test valoarea t_j:

$$P_{ij} = P(a >_C b \mid t(b)=t_i \text{ şi } t(a)=t_j) \quad (15.3.2.1)$$

$$P_{ij} = \frac{\sum_{k=2}^{m}\sum_{h=1}^{k-1} f_{j,k} f_{i,h}}{(\sum_{k=1}^{m} f_{j,k})(\sum_{h=1}^{m} f_{i,h})} \quad (15.3.2.2)$$

Se defineşte $P_{ji}=P_{ij}$, $i,j=1,2,...,n$.

Pentru fiecare valoare de test $t_i \in T$ se calculează probabilitatea ca doi subiecţi care au obţinut scorul t_i să aibă valori egale la variabila criteriu:

$$P_{ii} = P(a =_C b \mid t(a)=t_i \text{ şi } t(b)=t_i) \quad (15.3.2.3)$$

$$P_{ii} = \begin{cases} \dfrac{\sum_{k=1}^{m} f_{i,k}(f_{i,k}-1)}{(\sum_{k=1}^{m} f_{i,k})(\sum_{k=1}^{m} f_{i,k}-1)}, & daca \ \sum_{k=1}^{m} f_{i,k} > 1 \\ 1, & daca \ \sum_{k=1}^{m} f_{i,k} = 1 \end{cases} \quad (15.3.2.4)$$

Se obţine astfel o matrice simetrică $B=(P_{ij})_{i,j=1,...,n}$.

Probabilităţile P_{ij} care se abat cel mai mult de la valoarea 1 corespund perechilor de valori de test care nu pot fi interpretate corect. Atunci când $i<j$ şi $P_{ij}<0.5$, este foarte riscant să se afirme despre un subiect care a obţinut la test rezultatul t_j că este superior unei persoane care a realizat valoarea t_i.

Definim doi coeficienţi de validitate relativă la criteriu, prin următoarele formule (Albu, Piţariu, 1994):

$$V_{dif} = \min\{P_{ij} \mid i,j=1,2,...,n \text{ şi } i\neq j\} \quad (15.3.2.5)$$

$$V_{eg} = \min\{P_{ii} \mid i=1,2,...,n\} \quad (15.3.2.6)$$

Aceştia se interpretează astfel :

- Există o probabilitate de cel puţin $100V_{dif}$% ca un subiect a care a obţinut la test un rezultat mai bun decât un alt subiect b ($a>_T b$), să fie mai bun decât acesta şi în privinţa variabilei criteriu ($a>_C b$).
- Există o probabilitate de cel puţin $100V_{eg}$% ca doi subiecţi, a şi b, care au realizat aceeaşi valoare la test ($a=_T b$) să fie egali şi la variabila criteriu ($a=_C b$).

Legătura dintre validitatea relativă la criteriu a testului şi coeficienţii V_{eg} şi V_{dif} poate fi formulată astfel:

TEOREMA 15.3.2.1

a. Când $m=n$ atunci testul este valid dacă şi numai dacă $V_{eg}=V_{dif}=1$.
b. Dacă $m<n$ şi testul este valid atunci $V_{eg}=1$.
c. Dacă $m>n$ şi testul este valid atunci $V_{dif}=1$.

Se poate îmbunătăţi valoarea coeficientului V_{dif} pentru un test ţinând cont de următoarea observaţie: Dacă se înlocuiesc două valori de test consecutive, t_i şi t_{i+1}, cu o valoare nouă, $t_{i\&i+1}$, atunci fiecare probabilitate $P_{i\&i+1,k}$ va fi cuprinsă în intervalul

$$(\min\{P_{ik},P_{i+1,k}\}, \max\{P_{ik},P_{i+1,k}\}),$$

$k=1,2,...,n$.

Deci se poate mări valoarea coeficientului V_{dif} prin reunirea a două sau mai multe valori de test consecutive.

Înlocuind valorile t_i, t_{i+1},..., t_{i+k} , $k\geq 1$, din **T** cu o valoare nouă, t , cardinalul mulţimii **T**, n, scade. Coeficientul V_{dif} creşte, dar este posibil ca valoarea coeficientului V_{eg} să scadă.

Exemplu.

S-a aplicat acest procedeu pentru o scală nouă, **A**, alcătuită din itemi extraşi din Inventarul Psihologic California (CPI). Drept variabilă criteriu s-a folosit o scală existentă în CPI, scala Dominanţă (**Do**).

În figura 15.3.2.1 sunt reprezentate sub forma unei dendrograme grupările de scoruri realizate de algoritm.

În tabelul 15.3.2.1 sunt notate, pentru fiecare pas, valorile coeficienţilor V_{dif} şi V_{eg}. Se constată că, în timp ce coeficientul V_{dif} creşte de la 0.442, deci de la o valoare necorespunzătoare pentru test, la 0.793, coeficientul V_{eg} oscilează.

Se impune o observaţie: Coeficientul de corelaţie dintre scalele **A** şi **Do** este egal cu 0.856, pentru N=1065. În conformitate cu procedeul obişnuit de lucru, s-ar deduce că scala **A** are o validitate bună relativ la scala **Do**. Or, probabilitatea de a interpreta corect rezultatele obţinute de doi subiecţi este, pentru unele perechi de scoruri, mai mică decât 50%.

Tabelul 15.3.2.1

Valorile coeficienţilor de validitate
la fiecare pas al algoritmului

Grupele de scoruri formate	V_{dif}	V_{eg}
-	0.442	0.076
14,15	0.511	0.076
2,3	0.521	0.076
9,10	0.532	0.076
6,7	0.535	0.076
12,13	0.540	0.076
4,5	0.576	0.076
8,(9,10)	0.639	0.076
(12,13),(14,15)	0.645	0.076
0,1	0.664	0.082
(4,5),(6,7)	0.683	0.082
(0,1),(2,3)	0.689	0.079
(8,(9,10)),11	0.786	0.079
((12,13),(14,15)),16	0.793	0.079

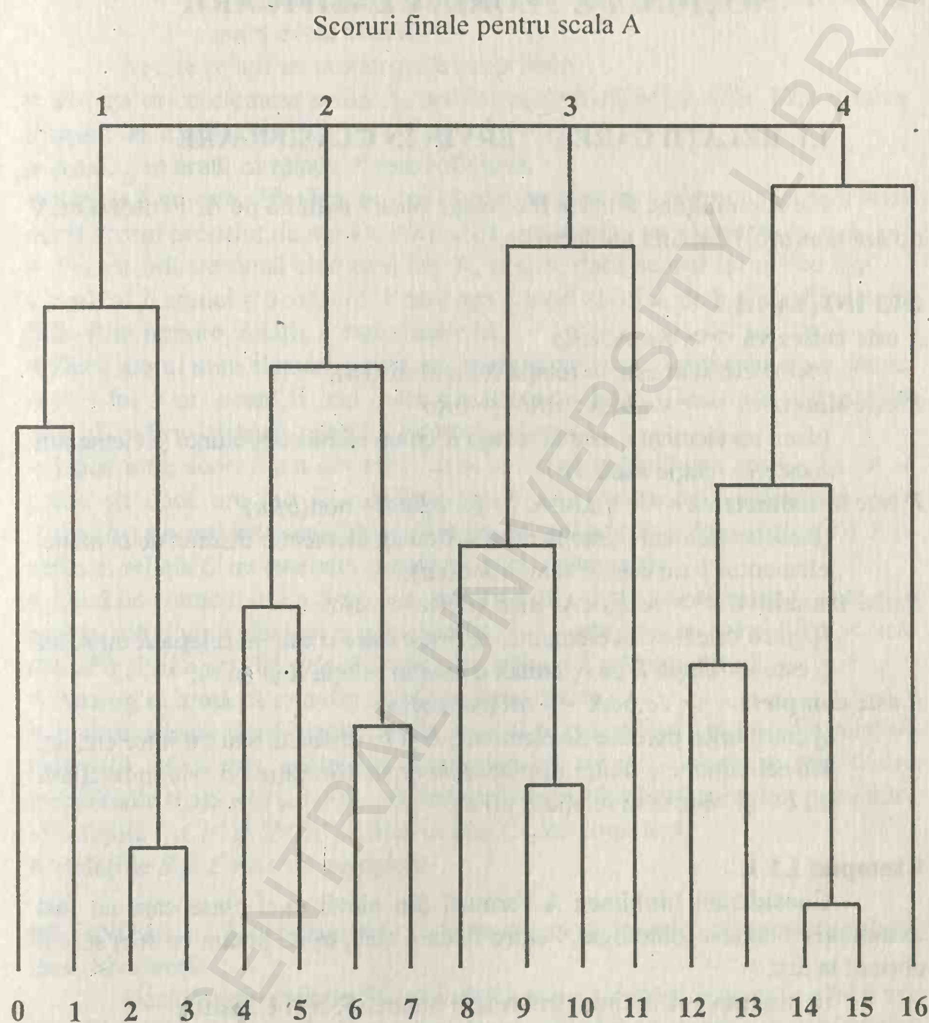


Figura 15.3.2.1
Gruparea scorurilor pentru scala A

ANEXA I
NOŢIUNI DE TEORIA CLASIFICĂRII

I.1. RELAŢII CARE INTERVIN ÎN CLASIFICARE

Fie A o mulţime finită şi R o relaţie binară definită pe A. Pentru $a,b \in A$ notăm non(aRb) dacă nu are loc aRb.

DEFINIŢIA I.1.1.

R este **reflexivă** $\Leftrightarrow \forall a \in A$: aRa

(orice element este în relaţia R cu el însuşi);

R este **simetrică** $\Leftrightarrow \forall a,b \in A$: $aRb \Rightarrow bRa$

(dacă un element a este în relaţia R cu un element b, atunci şi elementul b este în relaţia R cu a);

R este **antisimetrică** $\Leftrightarrow \forall a,b \in A$, $a \neq b$: $aRb \Rightarrow$ non(bRa)

(dacă un element a este în relaţia R cu un element b distinct de a, atunci elementul b nu este în relaţia R cu a);

R este **tranzitivă** $\Leftrightarrow \forall a,b,c \in A$: aRb şi $bRc \Rightarrow aRc$;

(pentru oricare trei elemente, a, b şi c dacă a este în relaţia R cu b, iar b este în relaţia R cu c, atunci a este în relaţia R şi cu c);

R este **completă** $\Leftrightarrow \forall a,b \in A \Rightarrow aRb$ sau bRa.

(pentru orice pereche de elemente, a şi b, distincte sau nu între ele, are loc cel puţin una dintre următoarele două afirmaţii: "a este în relaţia R cu b" şi "b este în relaţia R cu a).

Exemplul I.1.1.

Considerăm mulţimea A formată din elevii unei clase care au fost examinaţi cu un test psihologic. Pentru fiecare elev, $a \in A$, notăm cu $s(a)$ scorul obţinut la test.

În mulţimea A definim trei relaţii binare, S, E şi C, astfel: pentru oricare două elemente $a,b \in A$:

$aSb \Leftrightarrow s(a)>s(b)$

(a este în relaţia S cu b dacă a este "superior" lui b la test, adică a a obţinut un scor mai mare decât b);

$aEb \Leftrightarrow s(a)=s(b)$

(a este în relaţia E cu b dacă a este "egal" cu b la test, adică a şi b au obţinut acelaşi scor);

$aCb \Leftrightarrow s(a) \geq s(b)$

(*a* este în relaţia *C* cu *b* dacă *a* este "cel puţin la fel de bun ca şi *b* la test, adică *a* a obţinut un scor egal cu cel al lui *b* sau mai mare decât acesta).

Aceste relaţii au următoarele proprietăţi:

- Pentru orice element *a* din A, are loc $s(a)=s(a)$, adică *aEa*. Prin urmare, relaţia *E* este reflexivă.
- Analog se arată că relaţia *C* este reflexivă.
- Relaţia *S* nu este reflexivă, pentru că nici un elev nu a obţinut scor mai mare decât scorul propriu: nu are loc $s(a)>s(a)$, pentru nici un element *a* din A.
- Pentru oricare două elemente din A, *a* şi *b*, dacă scorul lui *a* este egal cu scorul lui *b* atunci şi scorul lui *b* este egal cu cel al lui *a*, deci, din *aEb* rezultă *bEa*. Prin urmare, relaţia *E* este simetrică.
- Dacă scorul unui element *a* este mai mare decât scorul unui element *b*, atunci scorul lui *b* nu poate fi mai mare decât scorul lui *a*. Deci, din *aSb* rezultă non(*bSa*). Prin urmare, relaţia *S* este antisimetrică.
- Dacă între scorurile a doi elevi, *a* şi *b*, există inegalitatea $s(a)\geq s(b)$, nu se poate şti dacă are loc şi inegalitatea $s(b)\geq s(a)$ (care ar fi adevărată dacă $s(a)=s(b)$) sau are loc non($s(b)\geq s(a)$) (ceea ce ar fi adevărat dacă $s(a)>s(b)$). Prin urmare, relaţia *C* nu este nici simetrică, nici antisimetrică.
- Dacă un element *a* din A are un scor egal cu cel al unui element *b*, care este egal cu cel al unui element *c*, atunci scorul lui *a* este egal cu cel al lui *c*. Adică, din *aEb* şi *bEc* rezultă *aEc*. Prin urmare, relaţia *E* este tranzitivă.
- Analog se arată că relaţiile *S* şi *C* sunt tranzitive.
- Pentru oricare două elemente din A, *a* şi *b*, este adevărată una din relaţiile $s(a)>s(b)$, $s(a)=s(b)$, $s(b)>s(a)$. Înseamnă că are loc cel puţin una dintre inegalităţile $s(a)\geq s(b)$ şi $s(b)\geq s(a)$, respectiv este adevărată cel puţin una dintre afirmaţiile "*aCb*" şi "*bCa*". Adică relaţia *C* este completă.
- Relaţiile *S* şi *E* nu sunt complete.

DEFINIŢIA I.1.2. O relaţie reflexivă, tranzitivă şi simetrică se numeşte **relaţie de echivalenţă.**

Dacă *R* este o relaţie de echivalenţă şi pentru două elemente $a,b\in$A are loc *aRb*, se spune că "*a* şi *b* sunt echivalente". Fiind dată o relaţie de echivalenţă pe mulţimea A, orice submulţime a lui A formată din elemente echivalente între ele formează o **clasă de echivalenţă.**

Exemplul I.1.2.

Relaţia *E* din exemplul I.1.1. este o relaţie de echivalenţă.

O clasă de echivalenţă este formată din toţi elevii care au obţinut acelaşi scor la test.

DEFINIŢIA I.1.3. O relaţie tranzitivă şi completă se numeşte **preordine totală.**

Exemplul I.1.3.

Relaţia C este o preordine totală.

Dacă R este o preordine totală şi pentru două elemente a şi b din mulţimea A are loc aRb vom spune că "a este preferat lui b".

DEFINIŢIA I.1.4. O preordine totală antisimetrică se numeşte **ordine totală**

Exemplul I.1.4.

Dacă în mulţimea A nu există elevi care să fi obţinut acelaşi scor, atunci relaţia S este o ordine totală.

TEOREMA I.1.1. Dacă R este o relaţie de preordine totală pe A, atunci relaţia I, definită pe A astfel:

pentru orice $a,b \in A$, $aIb \Leftrightarrow aRb$ şi bRa (I.1.1)

este o relaţie de echivalenţă.

Exemplul I.1.5.

Utilizând relaţiile definite în exemplul I.1.1 se constată că relaţia I definită prin formula (I.1.1) cu ajutorul relaţiei C

$aIb \Leftrightarrow aCb$ şi bCa

coincide cu relaţia E:

pentru orice $a,b \in A$, $aIb \Leftrightarrow s(a) \geq s(b)$ şi $s(b) \geq s(a) \Leftrightarrow s(a)=s(b) \Leftrightarrow aEb$

I.2. PARTIŢII

Fie E o mulţime finită cu n elemente.

DEFINIŢIA I.2.1. O **partiţie** a mulţimii E, P(E), este o mulţime de părţi ale lui E:

$P(E) = \{E_1, E_2, ..., E_k\}$, $E_i \subseteq E$, $i=1,2,...,k$

două câte două disjuncte:

$E_i \cap E_j = \emptyset$ pentru orice $i \neq j$ (I.2.1)

şi a căror reuniune este egală cu E:

$$E_1 \cup E_2 \cup ... \cup E_k = E \qquad (I.2.2)$$

Aceste părţi se numesc **clasele partiţiei.**

Indicarea unei partiţii pe o mulţime E este acelaşi lucru cu indicarea unei relaţii de echivalenţă pe E: două elemente $x,y \in E$ sunt echivalente dacă şi numai dacă x şi y aparţin aceleiaşi clase a partiţiei.

Dacă pe mulţimea E a fost definită o relaţie de preordine totală R, atunci relaţia de echivalenţă I definită prin formula (I.1.1) determină o partiţie a mulţimii E. Clasele partiţiei au proprietatea că ele pot fi aşezate în ordine, astfel încât, pentru orice clasă aleasă, E_i, fiecare element al ei, $a \in E_i$, este în relaţiile

$$aRb \text{ şi non}(bRa) \qquad (I.2.3)$$

cu fiecare element b al oricărei clase E_j aflate în dreapta clasei E_i.

Exemplul I.2.1.

Să reluăm exemplul I.1.1 şi să notăm cu $s_1, s_2,...,s_k$ scorurile care au fost obţinute de elevii mulţimii A, aşezate în ordine descrescătoare: $s_1 > s_2 > ... > s_k$.

Notăm cu A_i mulţimea elevilor din A care au obţinut scorul s_i, $i=1,2,...,k$.

Mulţimile $A_1, A_2,...,A_k$ formează o partiţie a mulţimii A. Orice element a al unei clase A_i este în relaţia aCb şi non(bCa), adică $s(a)>s(b)$, cu orice element b al claselor $A_{i+1}, A_{i+2},...,A_k$, pentru $i=1,2,...,k-1$.

DEFINIŢIA I.2.2. Un vector cu componente mulţimi, $(E_1, E_2, ..., E_k)$ $E_i \subseteq E$, $i=1,2,...,k$, este o **partiţie ordonată** pe mulţimea E, corespunzătoare relaţiei de preordine totală R dacă îndeplineşte următoarele condiţii:

- pentru orice mulţime E_i, $i=1,2,...,k$, şi oricare două elemente $a,b \in E_i$ au loc relaţiile aRb şi bRa;
- pentru oricare două mulţimi, E_i şi E_j, cu $i<j$, şi oricare două elemente $a \in E_i$ şi $b \in E_j$, au loc relaţiile aRb şi non(bRa).

Oricărei relaţii de preordine totală pe mulţimea E, R, îi corespunde o singură partiţie ordonată pe E.

Fie

$$(A_1, A_2,...,A_m) \qquad (I.2.4)$$

şi

$$(B_1, B_2,...,B_p) \qquad (I.2.5)$$

două partiţii ordonate pe mulţimea E, corespunzătoare la două relaţii de preordine totală, R_1 şi R_2.

DEFINIŢIA I.2.3. Se spune că partiţiile ordonate (I.2.4) şi (I.2.5) **coincid** dacă $m=p$ şi pentru orice $i=1,2,...,m$ are loc $A_i=B_i$.

DEFINIŢIA I.2.4. Se spune că partiţia ordonată (I.2.4) este o **rafinare a partiţiei ordonate** (I.2.5) dacă $m>p$ şi orice clasă de echivalenţă A_i, $i=1,2,...,m$, este o submulţime a unei clase de echivalenţă B_j, $j\in\{1,2,...,p\}$: $A_i\subseteq B_j$.

I.3. NUMEROTARE A UNEI PREORDINI TOTALE

Fie **E** o mulţime finită cu n elemente şi R o relaţie de preordine totală pe **E**.

DEFINIŢIA I.3.1. Se numeşte **numerotare** a elementelor lui **E** o funcţie bijectivă f: E -> $\{1,2,...,n\}$.

Prin urmare, numerotarea constă în atribuirea de numere consecutive elementelor din **E**, începând cu 1, astfel încât oricare două elemente să primească numere diferite.

Convenim să notăm cu e_i elementul din **E** pentru care $f(e_i)=i$, $i=1,2,...,n$.

DEFINIŢIA I.3.2. O numerotare a mulţimii **E**, f: **E** -> $\{1,2,...,n\}$ este o **numerotare a preordinii totale** R dacă pentru oricare două elemente distincte $a,b\in$E are loc implicaţia:

$$f(a)<f(b) \Rightarrow aRb \qquad \text{(I.3.1)}$$

Aceasta înseamnă că sunt adevărate relaţiile e_iRe_j pentru orice $i=1,2,...,n-1$ şi $j=i+1,i+2,...,n$.

Observaţie. Din faptul că pentru două elemente distincte $a,b\in$E are loc aRb nu se deduce $f(a)<f(b)$, întrucât este posibil ca împreună cu relaţia aRb să fie adevărată şi relaţia bRa.

TEOREMA I.3.1. Pentru orice relaţie de preordine totală R există cel puţin o numerotare a preordinii.

Demonstraţia acestei teoreme este constructivă, indicând metoda de definire a funcţiei f. Se procedează conform algoritmului următor:

(1) Se formează partiţia ordonată corespunzătoare relaţiei R pe mulţimea **E**:

$(E_1, E_2, ..., E_m)$

(2) Pentru fiecare clasă a partiţiei, E_i, i=1,2,...,m, se notează cu n_i numărul de elemente componente. Evident, $n_1+n_2+...+n_m=n$.

(3) Se defineşte o funcţie bijectivă f: $E \rightarrow \{1,2,...,n\}$ astfel:

$\forall e \in E_1$: $f(e) \in \{1,2,...,n_1\}$

$\forall e \in E_2$, $f(e) \in \{n_1+1, n_1+2, ..., n_1+n_2\}$

...

$\forall e \in E_m$, $f(e) \in \{n_1+n_2+...+n_{m-1}+1, n_1+n_2+...+n_{m-1}+2, ..., n_1+n_2+...+n_m\}$

Observaţie.

Cu excepţia cazului când fiecare clasă a partiţiei conţine un singur element, pentru preordinea totală R pot fi definite mai multe numerotări. Dar, orice numerotare f are proprietatea că media valorilor sale pe o clasă E_i este egală cu:

$$m(E_i) = n_{i-1} + \frac{n_i + 1}{2} \qquad \text{(I.3.2)}$$

i=1,2,...,m, cu convenţia n_0=0.

DEFINIŢIA I.3.3. Funcţia r: $E \rightarrow \mathbf{R}$ definită prin:

$\forall e \in E_i \quad r(e)=m(E_i), \quad i=1,2,...,m$ (I.3.3)

se numeşte **funcţie rang**. Valoarea ei pentru un element $e \in E$ se numeşte **rangul elementului** e în preordinea totală R.

Toate elementele din **E** care aparţin aceleiaşi clase de echivalenţă au acelaşi rang. Numărul valorilor distincte ale funcţiei rang este egal cu numărul claselor de echivalenţă ale partiţiei ordonate determinată de relaţia de preordine R.

Exemplul I.3.1.

Fie $E=\{e_1, e_2, e_3, e_4, e_5, e_6, e_7\}$ o mulţime de elevi care au fost examinaţi cu un test de atenţie. În următorul tabel este notat numărul greşelilor făcute de fiecare elev:

Elev	e_1	e_2	e_3	e_4	e_5	e_6	e_7
Nr. erori	10	11	9	10	8	12	11

Pe mulţimea **E** se defineşte o relaţie R, astfel:

$e_i R e_j \Leftrightarrow$ e_i a făcut mai puţine erori sau tot atâtea ca şi e_j, i=1,2,...,7.

Relaţia R este o relaţie de preordine totală pe mulţimea elevilor. Partiţia ordonată corespunzătoare ei este:
$(\{e_5\}, \{e_3\}, \{e_1, e_4\}, \{e_2, e_7\}, \{e_6\})$.

Valorile funcţiei rang r:**E** -> **R** se calculează astfel:

$r(e_5) = 0 + (1+1)/2 = 1$
$r(e_3) = 1 + (1+1)/2 = 2$
$r(e_1) = r(e_4) = 2 + (2+1)/2 = 3.5$
$r(e_2) = r(e_7) = 4 + (2+1)/2 = 5.5$
$r(e_6) = 6 + (1+1)/2 = 7$

I.4. CLASAMENTE

Fie **E** o mulţime finită cu n elemente, R o relaţie de preordine totală pe **E** şi r funcţia rang asociată relaţiei R.

DEFINIŢIA I.4.1. Se numeşte **clasament** sau **ierarhie** pe mulţimea E ordonarea elementelor mulţimii **E** stabilită de funcţia rang r, în care locul fiecărui element este egal cu rangul său.

Deci, un clasament este totdeauna asociat unei relaţii de preordine totală.

Exemplul I.4.1. Pentru elevii mulţimii **E** din exemplul I.3.1 clasamentul este următorul:

Elev	Loc
e_5	1
e_3	2
e_1	3.5
e_4	3.5
e_2	5.5
e_7	5.5
e_6	7

Să considerăm două relaţii de preordine totală pe mulţimea E, R_1 şi R_2, şi funcţiile rang asociate acestora, r_1 şi r_2.

Pentru fiecare element $e \in$**E** notăm cu $r_1(e)$ locul ocupat în primul clasament şi cu $r_2(e)$ locul ocupat în al doilea clasament.$_1r$ (e) şi$_2r$ (e) sunt rangurile elementului e în cele două preordini totale.

DEFINIŢIA I.4.2. Clasamentele asociate relaţiilor de preordine totală R_1 şi R_2 **coincid** dacă pentru orice element $e \in E$ are loc egalitatea $r_1(e)=r_2(e)$.

TEOREMA I.4.1. Clasamentele asociate relaţiilor de preordine totală R_1 şi R_2 coincid dacă şi numai dacă relaţiile R_1 şi R_2 sunt identice.

DEFINIŢIA I.4.3. Clasamentul asociat relaţiei de preordine totală R_1 este o **rafinare** a clasamentului asociat relaţiei de preordine totală R_2 dacă sunt îndeplinite următoarele condiţii:

- pentru oricare două elemente $a,b \in E$ are loc $r_1(a)<r_1(b)$ numai dacă $r_2(a)\leq r_2(b)$;
- pentru oricare două elemente $a,b \in E$ din egalitatea $r_1(a)=r_1(b)$ rezultă $r_2(a)=r_2(b)$.

Se constată că un clasamentul care este o rafinare a altui clasament are mai multe locuri distincte decât acesta.

TEOREMA I.4.2. Clasamentul asociat relaţiei de preordine totală R_1 este o rafinare a clasamentului asociat relaţiei de preordine totală R_2 dacă şi numai dacă partiţia ordonată corespunzătore relaţiei R_1 este o rafinare a partiţiei ordonate corespunzătoare relaţiei R_2.

Exemplul I.4.2.

Să reluăm exemplul I.3.1 şi să presupunem că la o altă administrare a testului numărul erorilor comise de elevi este cel prezentat în tabelul următor:

Elev	e_1	e_2	e_3	e_4	e_5	e_6	e_7
Nr. erori	5	7	4	6	2	10	9

Relaţia de preordine totală R_1 pe mulţimea elevilor fiind definită la fel ca în exemplul I.3.1, partiţia ordonată corespunzătoare ei este:
$(\{e_5\}, \{e_3\}, \{e_1\}, \{e_4\}, \{e_2\}, \{e_7\}, \{e_6\})$.

În acest caz, clasamentul elevilor este următorul:

Elev	Loc
e_5	1
e_3	2
e_1	3
e_4	4
e_2	5
e_7	6
e_6	7

Acest clasament este o rafinare a clasamentului corespunzător relaţiei de preordine totală *R* din exemplul I.3.1.

I.5. COMPARAREA A DOUĂ PARTIŢII

Fie **E** o mulţime şi

$$\{A_1, A_2, ..., A_n\}$$

şi

$$\{B_1, B_2, ..., B_m\}$$

două partiţii ale mulţimii **E**.

Pentru a vedea cât de asemănătoare sunt cele două partiţii, se construieşte un tablou de frecvenţe de următoarea formă:

Prima partiţie	A doua partiţie B_1 ... B_j ... B_m	
A_1 ... A_i ... A_n	$f_{1,1} \cdots f_{1,j} \cdots f_{1,m}$ ... $f_{i,1} \cdots f_{i,j} \cdots f_{i,m}$ ... $f_{n,1} \cdots f_{n,j} \cdots f_{n,m}$	$f_{1.}$ ... $f_{i.}$ ... $f_{n.}$
	$f_{.1} \ \dots \ f_{.j} \ \dots \ f_{.m}$	$f = N$

unde $f_{i,j}$ reprezintă numărul de obiecte care aparţin clasei A_i din prima partiţie şi clasei B_j din cea de-a doua partiţie, pentru i=1,2,...,n şi j=1,2,...,m.

Cele două partiţii concordă perfect dacă sunt identice, adică dacă ele conţin acelaşi număr de clase (m=n) şi fiecare clasă dintr-o partiţie este egală cu

o clasă din cealaltă partiţie. Aceasta înseamnă că în tabelul de frecvenţe de mai sus pe fiecare linie şi pe fiecare coloană există câte o singură frecvenţă nenulă.

O posibilitate de a cerceta dacă cele două partiţii au legătură între ele o oferă testul χ^2.

Se calculează valoarea variabilei χ^2 cu ajutorul formulei:

$$\chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{(f_{i,j} - t_{i,j})^2}{t_{i,j}} \tag{I.5.1}$$

unde, pentru $i=1,2,...,n$ şi $j=1,2,...,m$:
$f_{i,j}$ sunt frecvenţele observate, citite din tabel;
$t_{i,j}$ reprezintă frecvenţele teoretice, calculate cu formula:

$$t_{i,j} = \frac{f_{i.} f_{.j}}{f_{..}} \tag{I.5.2}$$

Se compară valoarea χ^2 astfel calculată cu valoarea $\chi^2_{(n-1)(m-1),\alpha}$ citită din tabelul legii χ^2 pentru pragul α fixat şi $(n-1)(m-1)$ grade de libertate.

Dacă $\chi^2 > \chi^2_{(n-1)(m-1),\alpha}$ atunci se respinge ipoteza că cele două partiţii sunt independente probabilistic între ele şi se acceptă ipoteza că între ele există legătură.

O altă metodă de comparare a partiţiilor este cea propusă de Huberty (1984), aplicabilă în cazul când $m=n$ şi fiecare clasă B_i corespunde clasei A_i, $i=1,2,...,n$. Procedeul său este util atunci când se consideră că partiţia $\{A_1, A_2,...,A_n\}$ este corectă şi se doreşte să se aprecieze calitatea celeilalte partiţii prin comparare cu aceasta. De exemplu, când se construieşte un test psihologic pentru determinarea tipului temperamental al subiecţilor, pentru a aprecia dacă testul funcţionează corect, se vor compara rezultatele furnizate de el cu cele date de un alt test care este apreciat ca fiind bun.

Metoda propusă de Huberty se bazează pe compararea frecvenţei relative a obiectelor plasate în aceeaşi clasă în ambele partiţii:

$$RS = \frac{\sum_{i=1}^{n} f_{i,i}}{f_{..}} \tag{I.5.3}$$

numită "raportul succesului", cu valoarea S care se consideră că s-ar obţine pentru acest raport dacă prima partiţie este corectă, iar a doua ar fi realizată la întâmplare.

Modul de calcul al valorii S depinde de ipotezele care se fac.

A. Se presupune că rapoartele $f_{i.}/f_{..}$, i=1,2,...,n, reflectă frecvenţele relative reale, în populaţia din care provine mulţimea de obiecte clasificate, ale situaţiilor cuprinse în clasele A_i, i=1,2,...,n.

În acest caz probabilitatea ca un obiect luat la întâmplare din populaţie să aparţină clasei A_i poate fi considerată egală cu $f_{i.}/f_{..}$, iar frecvenţa absolută a obiectelor din clasa A_i care vor fi repartizate în clasa B_i datorită hazardului va fi $f_i f_{i.}/f_{..}$, i=1,2,...,n. Pentru întreaga mulţime de obiecte clasificate frecvenţa relativă a concordanţelor datorate norocului va fi

$$S=\frac{\sum_{i=1}^{n} f_{i.}^2}{f_{..}^2} \qquad \text{(I.5.4)}$$

Dacă $RS \leq S$, atunci se poate aprecia că asemănările celor două partiţii sunt datorate întâmplării.

În caz contrar, se verifică ipoteza nulă ($RS<S$) aplicând un test statistic de comparare a unei frecvenţe relative (RS) cu o valoare dată (S).

B. Dacă frecvenţele relative $f_{i.}/f_{..}$, i=1,2,...,n, diferă de cele întâlnite în realitate şi una dintre ele este mult mai mare decât celelalte, atunci se calculează valoarea S pornind de la observaţia că s-ar putea obţine un număr mare de concordanţe repartizând toate obiectele în clasa A_i pentru care $f_{i.}/f_{..}$ are cea mai mare valoare. Deci se ia

$$S=\max\{f_{i.}/f_{..} | i=1,2,...,n\} \qquad \text{(I.5.5)}$$

şi se compară frecvenţa relativă RS cu S procedând ca în cazul precedent.

O altă metodă de comparare a două partiţii, aplicabilă în aceleaşi condiţii ca şi metoda lui Huberty, utilizează coeficientul K al lui Cohen (Bakeman, Gottman, 1986):

$$K=\frac{RS-S}{1-S} \qquad \text{(I.5.6)}$$

care compară raportul succesului *RS*, dat de formula I.5.3, cu valoarea

$$S = \sum_{i=1}^{n} \frac{f_{i.} f_{.i}}{f_{..}^2} \qquad (I.5.7)$$

Acest coeficient ia valori între -1 şi +1, concordanţa perfectă a celor două partiţii fiind indicată de valoarea maximă, +1.

I.6. COMPARAREA A DOUĂ CLASAMENTE

Compararea a două clasamente se bazează pe compararea partiţiilor ordonate corespunzătoare relaţiilor de preordine totală cărora le sunt asociate clasamentele.

Fie **E** o mulţime cu *n* elemente, R_1 şi R_2 două relaţii de preordine totală pe **E** şi

$(A_1, A_2,...,A_p)$ (I.6.1)

şi

$(B_1, B_2,...,B_m)$ (I.6.2)

partiţiile ordonate corespunzătoare lor.

Pentru a compara clasamentele asociate relaţiilor R_1 şi R_2 se compară partiţiile ordonate (I.6.1) şi (I.6.2).

În acest scop, se poate utiliza metoda lui Huberty şi coeficientul *K* al lui Cohen.

În plus, în lucrările de statistică sunt definiţi mai mulţi coeficienţi care permit compararea a două partiţii ordonate, şi care, având valori cuprinse între -1 şi 1 (sau între 0 şi 1) sunt uşor de interpretat.

Pentru fiecare element $e \in$ E, notăm cu $r_1(e)$, respectiv cu $r_2(e)$, rangul elementului *e* în cele două partiţii ordonate, (I.6.1) şi (I.6.2).

Unul dintre coeficienţii care servesc la compararea a două clasamente, cunoscut sub numele de **coeficientul de corelaţie a rangurilor al lui Spearman** este definit prin formula:

$$\rho = \frac{X_1 + X_2 - D}{2\sqrt{X_1 X_2}} \qquad (I.6.3)$$

unde:

$$D = \sum_{e \in E} \{r_1(e) - r_2(e)\}^2 \qquad (I.6.4)$$

$$X_i = \frac{n^3 - n}{12} - \frac{1}{12} \sum_{e \in E} \frac{1}{n_i(e)} \{n_i^3(e) - n_i(e)\} \qquad (I.6.5)$$

$n_i(e)$ reprezintă numărul de elemente care au acelaşi rang ca şi e în partiţia ordonată corespunzatoare relaţiei de preordine totală R_i, pentru i=1,2.

Dacă R_1 şi R_2 sunt relaţii de ordine totală, atunci fiecare clasă de echivalenţă din cele două partiţii ordonate conţine un singur element, prin urmare $n_1(e)=n_2(e)=1$ pentru orice element $e \in$E. Deci:

$$\rho = 1 - \frac{6D}{n^3 - n} \qquad (I.6.6)$$

Coeficientul ρ ia valoarea maximă, egală cu 1, dacă şi numai dacă partiţiile ordonate (I.6.1) şi (I.6.2) coincid. Dacă R_1 şi R_2 realizează clasamente "inverse" în sensul că $r_1(e)=n+1-r_2(e)$ pentru orice element $e \in$E, atunci ρ=-1.

Atunci când fiecare dintre partiţiile ordonate (I.6.1) şi (I.6.2) este asociată unei variabile măsurată pe scală de interval sau de raport, iar fiecare clasă de echivalenţă corespunde unei valori (măsuri) a acestei variabile, pentru compararea celor două partiţii ordonate se poate utiliza coeficientul de corelaţie liniară. În acest scop se întocmeşte un tabel de frecvenţe de următoarea formă:

Variabila X	Variabila Y $y_1 \dots y_j \dots y_m$	
x_1 ... x_i ... x_p	$f_{1,1} \cdots f_{1,j} \cdots f_{1,m}$ ... $f_{i,1} \cdots f_{i,j} \cdots f_{i,m}$ ... $f_{p,1} \cdots f_{p,j} \cdots f_{p,m}$	$f_{1.}$ ... $f_{i.}$ ... $f_{p.}$
	$f_{.1} \dots f_{.j} \dots f_{.m}$	$f_{..}=N$

unde $f_{i,j}$ reprezintă numărul de obiecte care au valoarea x_i pentru variabila X (aparţin clasei A_i din prima partiţie) şi valoarea y_j la variabila Y (aparţin clasei B_j din cea de-a doua partiţie), pentru i=1,2,...,p şi j=1,2,...,m.

Coeficientul de corelaţie liniară r între variabilele X şi Y se calculează cu formula:

$$r=\frac{\sum_{i=1}^{p}\sum_{j=1}^{m} x_i y_j f_{i,j} - Nm(X)m(y)}{N\sigma(X)\sigma(Y)} \tag{I.6.7}$$

unde:
$m(X)$ şi $m(Y)$ sunt mediile variabilelor X şi Y:

$$m(X)=\frac{\sum_{i=1}^{p} x_i f_{i.}}{N} \tag{I.6.8}$$

$$m(Y)=\frac{\sum_{j=1}^{m} y_j f_{.j}}{N} \tag{I.6.9}$$

$\sigma(X)$ şi $\sigma(Y)$ sunt abaterile standard ale variabilelor X şi Y:

$$\sigma(X)=\sqrt{\frac{\sum_{i=1}^{p}(x_i-m(X))^2 f_{i.}}{N-1}} \tag{I.6.10}$$

$$\sigma(Y)=\sqrt{\frac{\sum_{j=1}^{m}(y_j-m(Y))^2 f_{.j}}{N-1}} \tag{I.6.11}$$

Acest coeficient arată în ce măsură norul de puncte (x_i,y_j), $i=1,2,...,p$ şi $j=1,2,...,m$ se grupează în jurul unei drepte.

Când norul de puncte este format din puncte aşezate de-a lungul unei drepte, coeficientul de corelaţie liniară are valoarea +1 (dacă dreapta are pantă pozitivă) sau -1 (dacă dreapta are pantă negativă). Cu cât acest nor este mai

împrăştiat, cu atât valoarea coeficientului de corelaţie liniară este mai apropiată de 0.

Cu ajutorul testului Student (t) se poate verifica dacă valoarea unui coeficient de corelaţie liniară este semnificativ diferită de 0. Pentru aceasta se calculează valoarea t, cu ajutorul formulei:

$$t=r\sqrt{\frac{N-2}{1-r^2}} \qquad \text{(I.6.12)}$$

unde:
N este volumul mulţimii **E**;
r este valoarea calculată a coeficientului de corelaţie liniară.

Din tabelul legii Student se citeşte valoarea $t_{N-2,\alpha}$ corespunzătoare pragului de probabilitate α fixat şi la N-2 grade de libertate. Dacă $t>t_{N-2,\alpha}$ atunci se respinge ipoteza că valoarea coeficientului r este nesemnificativă statistic (nu diferă semnificativ de 0).

I.7. AGREGAREA CLASAMENTELOR

Prin **agregarea** mai multor clasamente într-unul singur se înţelege definirea unei ierarhii a elementelor unei mulţimi **E** plecând de la o mulţime dată de clasamente iniţiale.

Fie **E** o mulţime finită, m un număr natural, $m\geq 2$, şi S_E mulţimea relaţiilor de preordine totală existente pe **E**.

Notăm cu

$$S_E^m = \underbrace{S_E x S_E x \ldots x S_E}_{\text{de } m \text{ ori}}$$

mulţimea m-uplurilor de preordini totale ale elementelor din **E**.

DEFINIŢIA I.7.1. Se numeşte **procedeu de agregare** o funcţie

$$g:S_E^m\text{->}S_E.$$

Adică, un procedeu de agregare a clasamentelor construieşte un clasament pe mulţimea **E** cu ajutorul a m clasamente pe **E**.

Trebuie reţinut că nu orice funcţie care construieşte o relaţie binară completă R pe mulţimea **E** pornind de la m preordini totale pe **E** este un

procedeu de agregare, întrucât nu se asigură totdeauna tranzitivitatea relaţiei *R*. Un exemplu de acest fel îl constituie metoda deciziei majoritare (Arrow, 1963), care defineşte o relaţie *R* în mulţimea **E** utilizând *m* relaţii de preordine totală pe **E**, $R_1, R_2, \ldots, R_m$, astfel:

$$\forall a,b\in E : \quad aRb \iff \sum_{\substack{aR_ib \\ i=\overline{1,m}}} 1 \geq \sum_{\substack{bR_ia \\ i=\overline{1,m}}} 1 \qquad \text{(I.7.1)}$$

Adică *a* este preferat lui *b* în clasamentul final (are loc relaţia *aRb*) dacă numărul ierarhiilor în care *a* este preferat lui *b* nu este mai mic decât numărul ierarhiilor în care *b* este preferat lui *a*.

Este binecunoscut următorul exemplu, construit de marchizul de Condorcet în 1785 (Păun, 1987).

Se consideră trei candidaţi, *a*, *b* şi *c* şi 60 de alegători care propun, fiecare, câte o relaţie de ordine totală pe mulţimea {*a*,*b*,*c*} şi, deci, o ierarhie a elementelor acestei mulţimi în care nu pot să apară doi candidaţi pe acelaşi loc. Să presupunem că s-au obţinut următoarele ierarhii:

Locul			Frecvenţa
1	2	3	
a	b	c	de 23 de ori
b	c	a	de 17 ori
b	a	c	de 2 ori;
c	b	a	de 8 ori;
c	a	b	de 10 ori.

Să aplicăm metoda deciziei majoritare, folosind simbolul > pentru a arăta că un element se află înaintea altuia în ierarhie şi notând cu *R* relaţia binară construită:

$a>b$ de 33 de ori şi $b>a$ de 27 de ori => *aRb*
$b>c$ de 42 de ori şi $c>b$ de 18 ori => *bRc*
$a>c$ de 25 de ori şi $c>a$ de 35 de ori => *cRa*

Deci relaţia *R* nu este tranzitivă.

Două procedee de agregare, definite pentru cazul când ierarhiile de pornire corespund unor relaţii de ordine totală sunt metoda lui Borda (Păun, 1987) şi metoda lui Onicescu (1970).

Metoda lui Borda acordă m, m-1,..., 2, 1 puncte locurilor 1,2,...,m din ierarhiile de plecare, însumează punctele obţinute de fiecare element şi ordonează descrescător punctajele totale.

Metoda lui Onicescu este similară metodei anterioare, cu deosebirea că locurilor 1,2,...,m le acordă un număr de puncte:

$$1, \frac{1}{2}, \frac{1}{2^2}, \ldots, \frac{1}{2^{m-1}}$$

I.8. OPERAŢIILE CLASIFICĂRII

Clasificarea este o operaţie logică care acţionează asupra unei mulţimi de obiecte (sau fiinţe sau fenomene), fie pentru a le grupa, formând **clase**, fie pentru a le repartiza în clase existente.

Gruparea obiectelor în clase, prin clasificare, se face astfel încât gradul de asemănare (de similaritate) să fie mare între membrii aceleiaşi clase şi mic între membrii unor clase diferite.

În foarte multe cazuri clasele sunt disjuncte două câte două, determinând o *partiţie* a mulţimii obiectelor clasificate.

Alteori, prin clasificare se defineşte o *partiţie ordonată* şi o *ierarhie*. Această operaţie se numeşte **ierarhizare**. Clasele formate în acest caz sunt alcătuite din elemente echivalente între ele.

În timpul constituirii claselor se stabileşte şi o descriere a lor, sub forma prototipurilor. Un *prototip* este un obiect ideal al clasei (de exemplu, poate fi centrul de greutate al obiectelor grupate în clasă) sau o mulţime de obiecte ideale care îndeplinesc o anumită condiţie (de exemplu, valorile caracteristicilor lor verifică o ecuaţie). Termenul de "obiect ideal" desemnează un element care nu este necesar să aparţină mulţimii obiectelor clasificate, dar care este descris în acelaşi mod ca şi acestea, adică prin aceleaşi caracteristici.

Operaţia de determinare a clasei în care este cel mai potrivit să fie repartizat un obiect care nu aparţine mulţimii de obiecte pe baza cărora s-au format clasele este denumită **clasare** sau **identificare**. Ea acţionează de obicei prin compararea obiectului cu prototipurile claselor şi alegerea clasei al cărei prototip seamănă cel mai mult cu obiectul în cauză.

ANEXA II

CÂTEVA NOŢIUNI MATEMATICE ŞI STATISTICE CARE INTERVIN ÎN TEORIA TESTELOR PSIHOLOGICE

II.1 DEFINIŢIA VARIABILEI

O **variabilă** este o clasă generală sau o categorie de obiecte, evenimente, proprietăţi sau situaţii. Câteva exemple de variabile sunt: performanţa la un test de atenţie, inteligenţa, sexul, intensitatea zgomotului. În cadrul categoriei se pot observa diverse moduri de existenţă sau de manifestare a variabilei. De exemplu, pentru variabila "situaţia şcolară la sfârşitul anului şcolar, pentru elevi" sunt posibile cazurile "promovat", "corigent" şi "repetent". Acestea se numesc **niveluri** sau **valori** ale variabilei.

O variabilă care are numai două valori se numeşte **variabilă dihotomică** sau **variabilă booleană** sau **variabilă binară**. Aşa sunt, de exemplu, variabila "sex" şi variabila asociată unui item la care răspunsurile posibile sunt "da" şi "nu".

Unele variabile, precum sexul sau culoarea ochilor, au valori **calitative** (exprimate prin cuvinte), altele, cum sunt vârsta şi înălţimea, au valori **cantitative** (exprimate prin numere).

În cazul variabilelor cu valori numerice sunt posibile două situaţii:

(1) Variabila poate lua numai anumite valori numerice. Între oricare două valori ale variabilei există doar un număr fixat de valori posibile. Mulţimea tuturor valorilor variabilei este finită sau numărabilă. În acest caz se spune că variabila este **discretă**. De exemplu, numărul de copii al unei familii sau numărul de persoane dintr-o instituţie sunt variabile discrete.

(2) Variabila poate lua orice valoare numerică dintr-un interval finit sau infinit dat. Între oricare două valori ale variabilei există o infinitate de valori posibile. În acest caz se spune că variabila este **continuă**.

II.2. VARIABILE DEPENDENTE, VARIABILE INDEPENDENTE

Variabila **dependentă** este variabila ale cărei valori depind de o altă variabilă, numită variabilă **independentă**. La rândul ei, variabila independentă

poate să depindă de o altă variabilă, fiind, deci, variabilă dependentă. Relaţia de dependenţă se referă strict la două variabile.

Există două tipuri de variabile independente:

a. variabilele manipulate sunt cele pe care experimentatorul le controlează prin manipulare activă pentru a urmări schimbările măsurabile apărute la variabila dependentă;

b. variabilele nemanipulate sunt variabile de clasificare: în funcţie de valorile lor se grupează unităţile statistice şi apoi se verifică dacă între grupele formate există deosebiri în privinţa valorilor variabilei dependente.

Exemple.

1. În cercetările asupra învăţării, variabila independentă manipulată poate fi intervalul de timp în care este prezent un stimul, iar variabila dependentă este precizia cu care subiecţii recunosc stimulul. O altă variabilă independentă manipulată poate fi complexitatea stimulului, iar variabila dependentă, timpul necesar subiecţilor pentru recunoaşterea stimulului.

2. Se doreşte să se verifice ipoteza că deprinderile de rezolvare a problemelor morale depind de apartenenţa religioasă. În acest scop se formează grupe de persoane, în funcţie de religie, cărora li se administrează acelaşi test referitor la probleme morale. Se compară apoi rezultatele între grupe. În acest caz apartenenţa religioasă este variabila independentă nemanipulată.

Unii autori înţeleg prin variabilă independentă o variabilă care cauzează modificări ale variabilei dependente. Variabila dependentă este efect al variabilei independente. Dar, dacă nu se controlează prin manipulare variabila independentă, este dificil de tras concluzii asupra existenţei relaţiei cauză-efect între variabila independentă şi cea dependentă.

II.3. SCALE DE MĂSURĂ

Compararea a două persoane, *X* şi *Y*, conduce adesea la formularea unor întrebări de genul: Este *X* mai frumoasă decât *Y*? Cu cât este *X* mai înaltă decât *Y*? De câte ori este câştigul lui *Y* mai mare decât cel al lui *X*?

Descrierea unui obiect *X* poate fi făcută mai sugestivă raportând calităţile sale la cele ale populaţiei de obiecte din care face parte. Aceasta generează întrebări precum: Obiectul *X* este cu mult mai scump decât celelalte? Obiectul *X* este mai greu sau mai uşor decât celelalte? ş.a.m.d.

Formularea răspunsurilor la întrebări ca cele de mai sus este mult simplificată dacă se dau valori numerice caracteristicilor analizate, astfel încât

compararea obiectelor în privinţa fiecărei caracteristici să se reducă la compararea unor numere. Se realizează, în acest fel, *măsurarea* persoanelor sau a obiectelor.

Există mai multe definiţii ale măsurării. Majoritatea acestora sunt de forma următoare: "**măsurarea** constă în atribuirea de *x* lui *y* în conformitate cu *z*". În funcţie de autor:

x reprezintă "numere", "valori", "scoruri" sau "simboluri abstracte";

y este înlocuit cu "obiecte", "evenimente", "lucruri", "situaţii", "indivizi", "comportamente", "observaţii", "atribute", "proprietăţi" sau "răspunsuri";

z primeşte semnificaţia de "regulă specifică" sau "regulă de reprezentare" (Michell, 1997).

În cadrul acestei lucrări, prin măsurare a unei caracteristici la o mulţime de obiecte înţelegem atribuirea de numere obiectelor studiate, conform unei reguli bine precizate, prin care se urmăreşte ca anumite relaţii existente între obiecte, în privinţa caracteristicii respective, să existe şi între numerele atribuite lor. Mai precis, dacă în mulţimea obiectelor este definită o anumită relaţie (de exemplu, o relaţie de ordine), atunci trebuie să existe o relaţie similară în mulţimea măsurilor elementelor, iar dacă două obiecte sunt în relaţia considerată atunci măsurile lor trebuie să fie în relaţia similară ei. De pildă, dacă din punctul de vedere al caracteristicii cercetate un obiect *X* este superior unui obiect *Y*, atunci măsura obiectului *X* trebuie să fie mai mare decât măsura obiectului *Y*.

Înseamnă, deci, că măsurarea stabileşte un izomorfism între mulţimea obiectelor şi mulţimea măsurilor obiectelor. În asemenea condiţii se pot efectua diverse operaţii asupra măsurilor obiectelor, iar concluziile desprinse se extind asupra obiectelor.

Din punct de vedere matematic, măsurarea este o funcţie care ataşează fiecărui element din mulţimea de obiecte (mulţimea de definiţie a funcţiei), conform unor anumite reguli, un număr şi numai unul (din mulţimea în care funcţia ia valori), numit "măsura elementului".

Fiecare sistem de reguli impus defineşte câte un tip de măsurare sau câte un tip de **scală**. În general sunt utilizate patru tipuri de scale de măsură. În ordinea crescătoare a preciziei lor, acestea sunt:

- scala nominală;
- scala ordinală;
- scala de interval;
- scala de raport.

Regulile impuse pentru definirea unui tip de scală le includ, totdeauna, pe cele impuse pentru definirea scalelor aflate pe un nivel inferior. De exemplu,

dacă s-a putut defini o scală de interval, atunci s-ar fi putut defini şi o scală ordinală sau una nominală.

Trebuie reţinute două idei:

(1) Alegerea unui anumit nivel de măsurare pentru o mulţime de obiecte este dependentă de numărul şi tipul relaţiilor existente între obiecte şi care prezintă interes pentru cercetător. Evident, aceste relaţii se referă la caracteristica analizată.

(2) Prelucrările statistice care pot fi efectuate asupra măsurilor obiectelor pentru a cunoaşte mai bine obiectele sunt dependente de tipul de scală utilizat. Cu cât scala de măsură este mai precisă, cu atât prelucrările statistice permise sunt mai complexe şi conduc la concluzii mai numeroase.

Notăm în continuare cu **A** mulţimea obiectelor (sau a indivizilor sau a fenomenelor), cu **B** mulţimea de numere din care se dau valori măsurilor elementelor din **A** şi cu m:**A**->**B** funcţia care realizează măsurarea. Deci, pentru orice obiect $a \in$A, $m(a)$ reprezintă măsura sa.

Măsurarea pe o *scală nominală* poate fi realizată ori de câte ori caracteristica studiată permite împărţirea elementelor mulţimii **A** în clase disjuncte două câte două, adică atunci când există o **partiţie** a mulţimii **A**. În fiecare clasă sunt grupate toate obiectele care pot fi considerate echivalente în privinţa acestei caracteristici.

De pildă, variabilele referitoare la categoria socio-profesională sau la sexul persoanelor pot fi măsurate pe scală nominală.

Regula impusă la măsurarea nominală este ca toate elementele dintr-o aceeaşi clasă de echivalenţă să aibă aceeaşi măsură şi o valoare a măsurii să apară numai la elemente echivalente între ele. Adică:

R1. *Două elemente, $a,b \in A$, au măsuri egale ($m(a)=m(b)$) dacă şi numai dacă ele sunt echivalente ($a=b$).*

O posibilitate de definire a unei scale nominale constă în atribuirea de numere distincte claselor de obiecte. Toate obiectele unei clase vor avea aceeaşi măsură, egală cu numărul ataşat clasei. De exemplu, prin "numărarea" elementelor din **A** se realizează o măsurare nominală; în acest caz fiecare clasă este constituită dintr-un singur obiect. Numerele pe care le primesc jucătorii unei echipe de fotbal în cadrul unui joc sunt de asemenea măsuri nominale.

Măsura unui element este o "etichetă" atribuită elementului, care serveşte doar la denumirea, identificarea sau clasificarea acestuia. Ea este numerică, dar ar fi putut fi exprimată şi prin cuvinte. Din acest motiv nu are sens să se efectueze calcule aritmetice cu măsurile nominale.

Prin măsurare nominală se obţine o singură informaţie asupra elementelor mulţimii A: două elemente care au aceeaşi măsură sunt echivalente între ele. Se pot schimba între ele numerele atribuite claselor fără ca informaţia pe care ne-o furnizează o măsurare nominală să se schimbe.

Măsurarea pe o *scală ordinală* impune ca elementele mulţimii A să poată fi aranjate în ordine, dar cu permisiunea ca mai multe obiecte să ocupe acelaşi loc. De exemplu, muncitorii unei secţii pot fi ordonaţi ţinând cont de randamentul lor, de la cel mai eficient la cel mai puţin eficient.

Convenim să spunem că un element $a \in A$ este "superior" unui element $b \in A$ dacă în şirul ordonat format elementul a se află înaintea lui b. Notăm acest lucru prin $a>b$. Relaţia $>$ în cazul unei caracteristici concrete se poate traduce prin "este mai mare decât" sau "este mai bun decât" sau "este preferat lui" etc.

Dacă două elemente, $a,b \in A$, ocupă acelaşi loc în şirul ordonat, vom spune că a şi b sunt "echivalente" şi vom nota $a=b$. Relaţia $=$ poate însemna, în funcţie de context, "este la fel de mare ca" sau "este la fel de bun ca" sau "este apreciat la fel ca" etc.

Prin urmare, măsurarea pe scală ordinală poate fi realizată dacă în mulţimea A există o relaţie de preordine totală, pe care o notăm cu $\geq$. Oricare două elemente, a şi b din A, îndeplinesc una şi numai una dintre următoarele condiţii:

$a>b$, dacă $a \geq b$ şi non($b \geq a$);

$b>a$, dacă $b \geq a$ şi non($a \geq b$);

$a=b$, dacă $a \geq b$ şi $b \geq a$.

Măsurării ordinale i se impun **regulile** R1, de la scala nominală, şi:
R2. Pentru oricare două elemente $a,b \in A$ aflate în relaţia $a>b$ măsura elementului a este mai mare decât măsura elementului b: $m(a)>m(b)$.

Notăm cu $(A_1, A_2, ..., A_k)$ partiţia ordonată corespunzătoare preordinii totale $\geq$. Aici, pentru $i=1,2,...,k$, fiecare mulţime A_i conţine elemente aflate între ele în relaţia $=$ şi toate elementele din mulţimea A_i sunt în relaţia $>$ cu toate elementele mulţimilor $A_{i+1}, A_{i+2},...,A_k$.

Un algoritm simplu pentru definirea unei scale ordinale este următorul:

(1) se aleg k numere, $m_1>m_2>...>m_k$;

(2) tuturor elementelor din clasa A_i li se atribuie numărul m_i:

$$m(a)=m_i, \text{ pentru orice } a \in A_i,\ i=1,2,...,k.$$

Se realizează de asemenea o măsurare pe scală ordinală dacă fiecărui element a din A i se ataşează numărul $N-r(a)+1$, unde N este volumul mulţimii A, iar $r(a)$ este rangul elementului a în relaţia de preordine totală $\geq$.

Exemplu.

Să presupunem că mulţimea A={a,b,c,d,e,f,g} este formată din membrii unei echipe de handbal, care au primit după un meci următoarele calificative:

a : slab; *b* : foarte bun; *c* : slab;
d : foarte slab; *e* : mediocru; *f*: slab;
g : slab.

Pe baza calificativelor, în mulţimea jucătorilor se poate defini o relaţie de preordine totală, căreia îi corespunde partiţia ordonată ({b}, {e}, {a,c,f,g}, {d}). Se determină rangul fiecărui jucător şi măsurile calculate cu ajutorul rangului:

Jucător	Calificativ	Rang	Măsura
b	foarte bun	1	7
e	mediocru	2	6
a	slab	4.5	3.5
c	slab	4.5	3.5
f	slab	4.5	3.5
g	slab	4.5	3.5
d	foarte slab	7	1

Măsurarea ordinală dă următoarele informaţii:

- două elemente care au aceeaşi măsură sunt echivalente (nici una nu îi este superioară celeilalte);
- dintre două elemente cu măsuri neegale, cel care are măsura mai mare este superior celuilalt.

Trebuie reţinut că măsurarea ordinală nu spune nimic despre "distanţa" existentă între două elemente ale mulţimii A. Dacă $m(a)$-$m(b)$=$m(c)$-$m(d)$ nu înseamnă că *a* este tot cu atâta superior lui *b* cu cât *c* este superior lui *d*.

Transformând măsurile obiectelor din mulţimea A printr-o funcţie crescătoare, se obţine o altă măsurare ordinală a elementelor din A, întrucât noile măsuri păstrează ordinea obiectelor.

Măsurarea pe o *scală de interval* poate fi utilizată atunci când în mulţimea A există un clasament şi, în plus, se poate aprecia "distanţa" dintre oricare două elemente în privinţa caracteristicii studiate. Această "distanţă" este o funcţie

d : AxA -> R

care atribuie fiecărei perechi de elemente din A, (a,b), un număr nenegativ, $d(a,b)\geq 0$, şi îndeplineşte următoarele trei condiţii:

C1. "Distanţa" dintre oricare două elemente echivalente este egală cu zero: $a=b$ => $d(a,b)=0$.

C2. Pentru oricare două elemente $a,b \in$ A, "distanţa" dintre a şi b este egală cu "distanţa" dintre b şi a: $d(a,b)=d(b,a)$.

C3. Pentru oricare trei elemente $a,b,c \in$ A, între care există relaţiile ($a>b$ sau $a=b$) şi ($b>c$ sau $b=c$), are loc egalitatea:
$d(a,c) = d(a,b) + d(b,c)$

Unei scale de interval i se cere ca, pe lângă regulile R1 şi R2 impuse scalei ordinale, să îndeplinească şi următoarea **regulă**:
R3. Pentru oricare patru elemente $a,b,c,e \in A$ *are loc*
$|m(a)-m(b)| = |m(c)-m(e)|$ *dacă şi numai dacă* $d(a,b)=d(c,e)$.

Această regulă înseamnă că scala de interval atribuie unor obiecte egal distanţate între ele în privinţa caracteristicii cercetate, măsuri echidistante.

Un exemplu de măsurare pe scală de interval îl constituie evaluarea inteligenţei unor personalităţi, realizată de C. Cox pe baza parcurgerii unui material bibliografic vast. Au fost atribuite următoarele măsuri: 145 pentru Napoleon, 155 pentru Rembrandt, 165 pentru Beethoven şi 185 pentru Galileo Galilei. Prin aceşti coeficienţi de inteligenţă, Cox a vrut să afirme că "diferenţa dintre Rembrandt şi Napoleon este egală cu cea dintre Beethoven şi Rembrandt", iar "Galileo Galilei a fost mai inteligent decât Beethoven cu atâta cu cât Beethoven a fost mai inteligent decât Napoleon".

În cazul scalelor de interval, faptul că un element are măsura egală cu zero nu înseamnă că el este lipsit de caracteristica cercetată.

Exemplul tipic de scală de interval se întâlneşte la măsurarea temperaturii. După cum se ştie, temperatura este un parametru ce caracterizează starea de încălzire a unui sistem fizic şi se determină pe baza variaţiei unor mărimi fizice (volum, rezistenţă electrică, forţă electromotoare etc.), astfel încât unor variaţii egale ale mărimii fizice le corespund variaţii egale de temperatură. Termometrul cu care se înregistrează valorile temperaturii este gradat în unităţi echidistante. Indiferent ce reprezintă distanţa dintre două gradaţii succesive (un grad Celsius sau un grad Fahrenheit sau altceva) diferenţa de temperatură corespunzătoare distanţei dintre două gradaţii succesive este la fel de mare, fie că este vorba despre gradaţiile 2 şi 3 sau 10 şi 11. Se obişnuieşte să se atribuie unei temperaturi măsura zero, dar aceasta nu înseamnă absenţa totală a temperaturii. Zero grade Celsius corespunde, de exemplu, la 32 de grade Fahrenheit. Atât scala Celsius, cât şi scala Fahrenheit realizează măsurări de interval.

Un alt exemplu de măsurare pe scală de interval îl reprezintă calendarele. Unitatea de măsură pentru timp (an, lună, zi) este constantă în fiecare calendar. Dar, în calendarul evreiesc numărarea anilor se face începând cu "crearea lumii", a cărei dată este estimată prin secolul al IV-lea Î.C. Anului 1995 îi corespunde în calendarul evreiesc anul 5755.

Din aceste exemple se observă că măsurarea de tip interval fixează arbitrar punctul zero (mărimea caracteristicii căreia să i se atribuie măsura zero) şi lungimea unităţii de măsură ("distanţa" dintre două niveluri ale caracteristicii căreia să îi corespundă o diferenţă a măsurilor egală cu unu).

În cazul măsurării pe scală de interval, dacă măsurile obiectelor sunt transformate printr-o funcţie liniară:

$m_{nou}(a) = p + qm(a)$, pentru orice element $a \in A$

unde:

p şi q sunt numere reale, $q>0$,

$m(a)$ este măsura elementului a în scala de interval considerată,

$m_{nou}(a)$ este măsura nouă a elementului a

se obţine de asemenea o măsurare de interval.

Această proprietate face posibilă compararea între ele a mai multor caracteristici ale unui obiect, măsurate pe scală de interval, chiar dacă fiecare dintre măsurători furnizează valori din alt interval.

Exemplu.

Se administrează unei clase de elevi teste de cunoştinţe la gramatică, la matematică şi la biologie, fiecare test furnizând scoruri într-un alt interval. De pildă, la gramatică se pot obţine scoruri între 0 şi 20, la matematică între 0 şi 50, iar la biologie între 0 şi 100. Pentru a aprecia, la care dintre cele trei discipline este mai bun fiecare elev este necesar să se efectueze transformări prin care scorurile testelor să fie aduse într-un acelaşi interval de valori, de exemplu să fie cuprinse între -10 şi +10.

Măsurarea pe o *scală de raport* poate fi realizată ori de câte ori sunt îndeplinite condiţiile pentru definirea unei scale de interval şi, în plus, se poate aprecia în mod obiectiv când un element este lipsit de caracteristica cercetată.

De exemplu, se poate afirma cu siguranţă că un vehicul care staţionează este lipsit de viteză.

Scalei de raport i se impune ca, pe lângă regulile R1, R2 şi R3 să îndeplinească şi **regula** următoare:

R4. Un element $a \in A$ are măsura egală cu zero dacă şi numai dacă el este lipsit de caracteristica cercetată.

Exemplu.

A. Moles a realizat un clasament pentru 250 de compozitori cunoscuţi, în ordinea frecvenţei cu care au fost interpretate creaţiile lor în concerte, într-un interval de timp fixat. S-a constatat astfel că lucrările lui Schubert au fost incluse în 3.6% din programele de concert, cele ale lui Johann Strauss în 1.2%, iar cele ale lui Gershwin, în 0.3%. Deci, compoziţiile lui Schubert sunt interpretate de 3 ori mai frecvent decât cele ale lui Johann Strauss şi de 12 ori mai frecvent decât cele ale lui Gershwin. Variabila a cărei valoare pentru fiecare compozitor este frecvenţa cu care sunt interpretate creaţiile sale, exprimată în procente sau în număr de cazuri, este măsurată pe scală de raport. Ea are valoarea 0 doar dacă nici una dintre compoziţiile persoanei nu a fost inclusă în programele de concert.

Atunci când caracteristica măsurată reprezintă spaţiu (lungime, suprafaţă, volum), masă sau timp trecut de la un eveniment fixat (de exemplu, durată de viaţă, timp de reacţie) se poate stabili în mod obiectiv ce înseamnă "absenţa caracteristicii". Pentru asemenea variabile se pot defini măsurări de raport.

Deosebirea dintre o măsurare de interval şi una de raport constă în faptul că, la măsurarea de interval, din egalitatea $m(a) = Km(b)$ nu se poate trage concluzia că a este "de K ori mai bun" decât b (în sensul relaţiei >), dar acest lucru este posibil la măsurarea de raport. De pildă, indiferent dacă înălţimea persoanelor este măsurată în mm, m sau inch, un copil care are 160 cm este de două ori mai înalt decât unul care are 80 cm. Dacă o maşină se deplasează cu 60 km/oră iar alta cu 30 km/oră, se poate spune că prima merge de două ori mai repede decât a doua.

În cazul măsurării pe scală de raport, transformând măsurile obiectelor printr-o funcţie liniară:

$m_{nou}(a) = p + qm(a)$, pentru orice element $a \in A$

cu $p, q \in \mathbf{R}$ şi $q>0$, dacă numărul p este diferit de zero nu se va mai obţine o măsurare pe scală de raport, ci doar o măsurare pe scală de interval. De această observaţie trebuie să se ţină seama atunci când se efectuează transformarea valorilor unei variabile în cote standard.

Gruparea valorilor unei variabile măsurate pe scală de interval sau de raport poate conduce la transformarea acesteia într-o variabilă măsurată pe scală ordinală.

Exemplu.

Variabila "vârstă", care are ca valoare, pentru fiecare persoană, numărul de ani împliniţi, este măsurată pe scală de raport. Dacă se stabilesc trei grupe:

"sub 25 de ani", "între 25 şi 60 de ani" şi "peste 60 de ani", cărora li se atribuie valorile 1, 2 şi 3, atunci variabila va fi măsurată pe scală ordinală.

Scala de interval şi cea de raport sunt denumite **scale metrice**. Variabilele ale căror valori se obţin printr-o măsurare pe scală de interval sau pe scală de raport se numesc **variabile de tip scor**.

Alegerea unui tip de scală pentru măsurarea unei variabile trebuie să aibă în vedere, în primul rând, tipul de prelucrări statistice care ar trebui efectuate asupra variabilei. Dacă relaţiile cunoscute între elementele mulţimii **A** nu permit măsurarea variabilei pe scala care ar fi necesară calculelor statistice respective, se va realiza o măsurare pe o scală inferioară dar cât mai apropiată de aceasta.

Atunci când "obiectele" care se măsoară sunt comportamente, apare o problemă suplimentară, cunoscută sub denumirea de "reactivitate". Se spune că o măsură este **reactivă** dacă operaţia de măsurare schimbă comportamentul subiecţilor (Cozby, 1993). Când persoanele se ştiu observate, comportamentul lor poate să difere de cel pe care l-ar avea în condiţii normale. Chiar şi parametrii fiziologici îşi pot schimba valorile din cauza prezenţei echipamentelor de înregistrare a lor. Din acest motiv, se recomandă ca înainte de a efectua măsurătorile să i se lase timp subiectului să se obişnuiască cu prezenţa observatorului sau a aparatului de măsură.

II.4. DISTANŢE

Fie **E** o mulţime de vectori din spaţiul euclidian M-dimensional ($M \geq 2$) şi $x,y \in$ **E** două elemente oarecare:

$x=(x_1,x_2,\ldots,x_M)$

$y=(y_1,y_2,\ldots,y_M)$.

De exemplu, **E** poate corespunde unei mulţimi de persoane care au răspuns la M teste psihologice. Pentru fiecare persoană există în **E** un vector, care are ca şi componente scorurile realizate la teste, ordinea testelor fiind fixată (adică, cea de-a i-a componentă va conţine scorul la testul i).

Pentru calculul distanţei dintre x şi y se pot folosi următoarele norme ale spaţiului $\mathbf{R}^M$:

- distanţa Manhattan sau city block:

$$D_1(x,y)=\sum_{i=1}^{M}|x_i-y_i| \qquad \text{(II.4.1)}$$

• distanţa euclidiană:

$$D_2(x,y)=\sqrt{\sum_{i=1}^{M}(x_i-y_i)^2} \qquad \text{(II.4.2)}$$

• distanţa valorii maxime:

$$D_{max}(x,y)=\max\{|x_i-y_i| : i=1,2,...,M\} \qquad \text{(II.4.3)}$$

• distanţa Hamming, aplicabilă atunci când fiecare componentă a vectorilor are doar un număr finit de valori: distanţa Hamming dintre vectorii x şi y este dată de numărul componentelor în care cei doi vectori diferă.

• distanţa lui Mahalanobis:

$$D_M(x,y)=\{(x-y)^T C^{-1}(x-y)\}^{1/2} \qquad \text{(II.4.4)}$$

unde:

$(x-y)^T$ reprezintă transpusa vectorului coloană $x-y$;

$C=(\sigma_{ij})_{i,j=1,2,...,M}$ este matricea de covarianţă a celor M caracteristici pe mulţimea **E**.

Atunci când componentele vectorilor din mulţimea **E** sunt valorile a M variabile şi aceste variabile sunt necorelate liniar două câte două pe mulţimea E, are loc $\sigma_{ij}=0$ pentru $i,j=1,2,...,M,\ i\neq j$; deci matricea C^{-1} va fi o matrice diagonală, $C^{-1}=(c_{ij})_{i,j=1,2,...,M}$, cu $c_{ij}=0$ şi $c_{ii}=1/\sigma_{ii}$, $i,j=1,2,...,M,\ i\neq j$.

Observaţii

1. În cazul distanţei euclidiene şi a distanţei city block se poate obţine aceeaşi valoare a distanţei dintre doi vectori în două situaţii foarte deosebite: când diferă câte puţin valorile fiecărei componente şi când diferă mult valorile unei singure componente, celelalte fiind egale. Din acest motiv ele nu sunt potrivite să fie folosite atunci când o diferenţă mare între valorile unei singure componente (indiferent care) trebuie să conducă la concluzia că obiectele se deosebesc mult între ele. În asemenea situaţii este mai potrivită distanţa valorii maxime.

Exemplu.

Distanţele dintre vectorii:

$x=(0,0,0,0,0,0,0,0,0)$

$y=(1,1,1,1,1,1,1,1,1)$

$z=(3,0,0,0,0,0,0,0,0)$

$v=(9,0,0,0,0,0,0,0,0)$

sunt: $D_1(x,y)=D_1(x,v)=9$, $D_1(x,z)=3$

$D_2(x,y)=D_2(x,z)=3$, $D_2(x,v)=9$

$D_{max}(x,y)=1$, $D_{max}(x,z)=3$, $D_{max}(x,v)=9$

2. Distanţa euclidiană, spre deosebire de distanţa city block, "amplifică" distanţele dintre componente, făcând ca distanţele mari să pară şi mai mari (pentru că în formula de calcul a distanţei euclidiene diferenţele valorilor componentelor se ridică la pătrat). Adică, pentru distanţa euclidiană, o diferenţă mare între valorile unei componente contează mai mult decât diferenţele mici sau medii între valorile a două componente.

Exemplu

Distanţele dintre vectorii

$x = (0,0,0,0,0)$
$y = (1,1,1,1,0)$
$z = (1,1,2,0,0)$

sunt:

$D_1(x,y) = D_1(x,z)=4$
$D_2(x,y) = 2 \quad D_2(x,z) = 2.449$

3. Distanţa Hamming este potrivită atunci când nu are importanţă cât de mult se deosebesc între ele valorile componentelor, ci numai dacă ele sunt diferite.

Exemplu.

Distanţele Hamming dintre vectorii:

$x=(0,0,0,0,0,0,0,0,0)$
$y=(1,1,1,1,1,1,1,1,1)$
$z=(3,0,0,0,0,0,0,0,0)$
$v=(9,0,0,0,0,0,0,0,0)$

sunt:

$D_H(x,y)=9, \quad D_H(x,z)=D_H(x,v)=1$

4. Calculul distanţei lui Mahalanobis ţine cont şi de felul în care valorile componentelor covariază (variază unele în raport cu altele), dar necesită cunoaşterea tuturor vectorilor din mulţimea **E** şi calculul covarianţelor componentelor.

II.5. COTE STANDARD, COTE STANDARDIZATE

Fie X=$\{x_1,x_2,\ldots,x_n\}$ o mulţime de obiecte şi V o variabilă care a fost măsurată, folosind o scală de interval sau de raport, la toate elementele mulţimii X.

Notăm:

m = media valorilor variabilei V pe mulţimea X:

$$m=\frac{\sum_{i=1}^{n} x_i}{n} \qquad (II.5.1)$$

σ = abaterea standard a valorilor variabilei *V* pe mulţimea **X**:

$$\sigma=\sqrt{\frac{\sum_{i=1}^{n}(x_i-m)^2}{n}} \qquad (II.5.2)$$

Cota standard *z* corespunzătoare unei valori x_i din **X** se calculează cu formula:

$$z_i=\frac{x_i-m}{\sigma} \qquad (II.5.3)$$

Cota standard z_i arată cu câte unităţi de abatere standard se distanţează valoarea x_i de media valorilor variabilei *V*, *m*.

Transformarea în cote standard nu modifică proporţionalitatea intervalelor scalei. Astfel, dacă x_i, x_j şi x_k reprezintă trei valori ale variabilei *V*, iar z_i, z_j

$$\frac{x_i-x_j}{x_j-x_k}=\frac{z_i-z_j}{z_j-z_k} \qquad (II.5.4)$$

şi z_k sunt cotele standard corespunzătoare, are loc egalitatea:
care arată că distanţele relative între valorile variabilei *V* rămân aceleaşi.

Inconvenientul utilizării cotelor *z* îl constituie faptul că acestea se exprimă prin numere cu semn şi cu zecimale, fiind cuprinse, în general, în intervalul [-3, +3]. O rezolvare practică o constituie efectuarea unei transformări liniare asupra cotelor *z*, care conduce la obţinerea unor **cote standardizate**:

$$z' = M + Sz \qquad (II.5.5)$$

unde:

z' = cota transformată a cotei standard z;

M,S = media şi abaterea standard pentru cotele standardizate obţinute prin transformare.

În mod obişnuit, valorile z' se rotunjesc la numere întregi.

În tabelul II.5.1 sunt prezentate câteva dintre transformările utilizate mai des (Lohse, Ludwig, Röhr, 1982).

Tabelul II.5.1

Valorile M şi S pentru diverse transformări ale cotelor standard

	Scala							
	N	C	WP	L	T	Z	IQ	H
M	3	5	10	10	50	100	100	50
S	1	2	3	5	10	10	15	14

Cotele standardizate au aceleaşi proprietăţi ca şi cotele standard z, singura diferenţă între ele fiind valoarea mediei şi cea a abaterii standard.

BIBLIOGRAFIE

Aiken, L.R. (1994), *Psychological Testing and Assessment*, Allyn and Bacon, Massachusetts.

Albu, M. (1990), *Formule pentru estimarea fidelităţii testelor psihologice*, "Studia Univ. Babeş-Bolyai", Psychologia-Paedagogia, 2, pag. 84-98.

Albu, M. (1992), *O nouă metodă de interpretare a scorurilor obţinute la scalele chestionarelor de personalitate*, "Studia Univ. Babeş-Bolyai", Psychologia-Paedagogia, 1-2, pag. 141-143.

Albu, M. (1993), *Analiza de itemi: o etapă obligatorie în construirea unui test psihologic*, "Revista de psihologie", 1, pag. 75-81.

Albu, M. (1994), *O metodă nouă de interpretare a scorurilor la scalele Inventarului Psihologic California (CPI)*, "Studia Univ. Babeş-Bolyai", Psychologia-Paedagogia, 1-2, pag. 137-141.

Albu, M., Pitariu, H. (1991a), *O reconfigurare a noţiunii de putere de discriminare a testelor psihologice şi de cunoştinţe*, "Studia Univ. Babeş-Bolyai", Psychologia-Paedagogia, 1, pag. 87-92.

Albu, M., Pitariu, H. (1991b), *Algoritm de construire a unei scale pentru un test psihologic: contribuţii la reproiectarea scalei F/M a Inventarului Psihologic California (CPI)*, "Studia Univ. Babeş-Bolyai", Psychologia-Paedagogia, 2, pag. 29-35.

Albu, M., Pitariu, H. (1992a), *Optimizarea construcţiei testelor psihologice şi de cunoştinţe*, "Revista de psihologie", 4, pag. 339-344.

Albu, M., Pitariu, H. (1992b), *Un algoritm de construire a unei scale cu o consistenţă internă ridicată*, "Studia Univ. Babeş-Bolyai", Psychologia-Paedagogia, 1-2, pag. 141-143.

Albu, M., Pitariu, H. (1993), *Proiectarea testelor de cunoştinţe şi examenul asistat pe calculator*, Editura Casa Cărţii de Stiinţă, Cluj-Napoca.

Albu, M., Pitariu, H. (1994), *Redefining Criterion Validity of Psychological Tests*, "Revue Roumaine de Psychologie", 1, pag. 79-87.

Albu, M., Pitariu, H. (1995a), *Criterion-Related Validity of Psychological Tests: A New Solution*, "European Journal of Psychological Assessment", Supplement 1, pag. 56.

Albu, M., Pitariu, H. (1995b), *A New Method for Assessing the Stability over Time of Psychological Test Scores*, "Studia Univ. Babeş-Bolyai", Psychologia-Paedagogia, 1-2, pag. 35-44.

Albu, M., Pitariu, H. (1995c), *O contribuţie metodologică la formarea grupelor contrastante*, În: *Studii şi cercetări din domeniul ştiinţelor socio-umane, prezentate la sesiunile ştiinţifice anuale 1993-1994*, Academia Română - Filiala Cluj-Napoca, Institutul de Cercetări Socio-Umane, Cluj-Napoca.

Amelang, M., Borkenau, P. (1986), *The Trait Concept: Current Theoretical Considerations, Empirical Facts, and Implications for Personality Inventory Constructions*, În: Angleitner, A., Wiggins, J.S. (Ed.), *Personality Assessment via Questionnaires*, Springer-Verlag, Berlin, Heidelberg.

Anastasi, A. (1954 şi 1976), *Psychological Testing*, MacMillan Publishing Co., Inc., New York.

Andrich, D. (1988), *Rasch Models for Measurement*, Sage Publications, Inc.

Angleitner, A., John, O.P., Löhr, F.L. (1986), *It's What You Ask and How You Ask It: An Itemmetric Analysis of Personality Questionnaires*, În: Angleitner, A., Wiggins, J.S. (ed.), *Personality Assessment via Questionnaires*, Springer-Verlag, Berlin, Heidelberg.

Arrow, K. J. (1963), *Social Choice and Individual Values*, New Haven and London, Yale University Press.

Ausubel, D. P., Robinson, F. G. (1981), *Învăţarea în şcoală. Introducere în psihologia pedagogică*, Editura Didactică şi Pedagogică, Bucureşti.

Averill, J.R. (1980), *The Emotions*. În: Staub, E. (ed.), *Personality. Basic Aspects and Current Research*, Prentice-Hall, Englewood Cliffs, New Jersey.

Bakeman, R., Gottman, J. M. (1986), *Observing Interaction. An Introduction to Sequential Analisys*, Cambridge University Press.

Barclay, J.R. (1968), *Controversial Issues in Testing*, Houghton Mifflin Company, Boston.

Baron, R.M., Kenny, D.A. (1986), *The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations*, "Journal of Personality and Social Psychology", 6, vol. 51, pag. 1173-1182.

Benziman, H., Toder, A. (1993), *The Psychodiagnostic Experience: Reports of Subjects*, În: Nevo, B., Jäger, R. S. (ed.), *Educational and Psychological Testing: The Test Taker's Outlook*, Hogrefe & Huber Publishers.

Blum, M. L., Naylor, J. C. (1968), *Industrial Psychology. Its Theoretical and Social Foundations*, Harper & Row Publishers.

Boekaerts, M. (1996), *Personality and the Psychology of Learning*, "European Journal of Personality", vol. 10, 377-404.

Brogden, H. E., Taylor, E. K., (1949), *The Dollar Crtiterion - Applying the Cost Accounting Concept to Criterion Construction*, Personnal Psychology.

Burisch, M. (1986), *Methods of Personality Inventory Development- A Comparative Analysis*, În: Angleitner, A., Wiggins, J.S. (Ed.), *Personality Assessment via Questionnaires*, Springer-Verlag, Berlin, Heidelberg.

Burisch, M. (1997), *Test Length and Validity Revisited*, "European Journal of Personality", vol. 11, pag. 303-315.

Cascio, W. F. (1991), *Applied Psychology in Personnel Management*, Prentice-Hall, Inc., New Jersey.

Chaplin, W.F., John, O.P., Goldberg, L.R. (1988), *Conceptions of States and Traits: Dimensional Attributes with Ideals as Prototypes*, "Journal of Personality and Social Psychology", 4, vol. 54, pag. 541-557.

Cone, D. J. (1993), *The Current State of Behavioral Assessment*, "European Journal of Psychological Assessment", vol. 9, 3, pag. 175-181.

Corsini, R. J. (ed.) (1994), *Encyclopedia of Psychology*, John Wiley & Sons.

Cozby, P.C. (1993), *Methods in Behavioral Research*, Mayfield Publishing Company.

Crehan, K. D. (1974), *Item Analysis for Teacher-made Mastery Tests*, "Journal of Educational Measurement", vol. 11, 4, pag. 255-262.

Cresin, R. (1966), *Aplicarea indicatorului "energie informaţională" în cercetarea longevităţii*, "Revista de statistică", 12.

Cronbach, L. J. (1943), *On Estimates of Test Reliability*, "The Journal of Educational Psychology", 34, pag. 485-494.

Cronbach, L. J. (1966), *Essentials of Psychological Testing*, A Harper International Edition.

Cureton, E. (1955), *Validity*, În: Lindquist, E. F. (ed.), *Educational Measurement*, American Council on Education, Washington, D.C.

Dahlstrom, W.G. (1993), *Small Samples, Large Consequences*, "American Psychologist", vol. 48, nr. 4, pag. 393-399.

Davidsen, O. M., Maxey, J. (1993), *Correlates of Students' Reaction to their Testing Environment*, În: Nevo, B., Jäger, R. S. (ed.), *Educational and Psychological Testing: The Test Taker's Outlook*, Hogrefe & Huber Publishers.

Davis, F. B. (1955), *Item Selection Techniques*, În: Lindquist, E. F. (ed.), *Educational Measurement*, American Council on Education, Washington, D.C.

De Gruijter, D. N. M., Van der Kamp, L. J. Th. (1984), *Statistical Models in Psychological and Educational Testing*, Swets & Zeitlinger, Lisse.

Deinzer, R., Steyer, R., Eid, M., Notz, P. ş.a. (1995), *Situational Effects in Trait Assessment: the FPI, NEOFFI, and EPI Questionnaires*, "European Journal of Personality", vol. 9, pag. 1-23.

Downs, S. (1989), *Job Sample and Trainability Tests*, în: Herriot, P. (ed.), *Assessment and Selection in Organizations. Methods and Practice for Recruitment and Appraisal*, John Wiley & Sons, Chichester.

Ebel, R. L. (1955), *Writing the Test Item*, În: Lindquist, E. F. (ed.), *Educational Measurement*, American Council on Education, Washington, D.C.

Edwards, A. L. (1970), *The Measurement of Personality Traits by Scales and Inventories*, Holt, Rinehart and Winston, Inc.

Endler, N. S., Magnusson, D. (1976), *Interactional Psychology and Personality*, Hemisphere Publishing Corporation, Washington, D.C.

English, H. B., English, A. C. (1970), *A Comprehensive Dictionary of Psychological and Psychoanalytical Terms*, David McKay Company, Inc.

Everitt, B. S., Dunn, G. (1983), *Advanced Methods of Data Exploration and Modelling*, Heinemann Educational Books, London, Exeter.

Eyde, L.R., Kowal, D.M. (1987), *Computerised Test Interpretation Services: Ethical and Professional Concerns Regarding U.S. Producers and Users*, "Applied Psychology: An International Review", 36 (3/4), pag. 401-417.

Fernández-Ballesteros, R. (1993), *Behavioral Assessment: Dying, Vanishing or Still Running?*, "European Journal of Psychological Assessment", vol. 9, nr. 3, pag. 159-174.

Fife-Schaw, C. (1995), *Questionnaire Design*, În: Breakwell, G. M., Hammond, S., Fife-Schaw, C. (ed.), *Research Methods in Psychology*, Sage Publications.

Fisseni, H. J. (1990), *Lehrbuch der psychologischen Diagnostik, Verlag für Psychologie*, Dr. C. J. Hogrefe, Göttingen.

Flanagan, J. C. (1955), *Units, Scores and Norms*, În: Lindquist, E. F. (ed.), *Educational Measurement*, American Council on Education, Washington, D. C.

Fletcher, C. (1997), *The Future of Psychometric Assessment: Fostering Positive Candidate Attitude and Reactions*, "Selection & Development Review", vol. 13, nr. 4, pag. 8-11.

Funder, D.C. (1983), *Three Issues in Predicting More of the People: A Reply to Mischel and Peake*, "Psychological Review", vol. 90, nr. 3, pag. 283-289.

von Gilmer, H. B. (1971), *Industrial and Organizational Psychology*, McGraw-Hill Book Company.

Goldberg, L.R. (1996), *A Public-Domain, Broad-Bandwidth, Personality Inventory Measuring the Lower-Level Facets of Several Five-Factor Models*, Comunicare prezentată la a VIII-a Conferinţă Europeană asupra Personalităţii, Ghent.

Gough, H. G. (1989), *The New California Psychological Inventory: Implications for Research and Practice*, Lucrare dată participanţilor la simpozionul "Recent Advances in Psychological Assessment", Nevada, 29 aprilie 1989.

Gough, H. G. (1991), *Scales and Combinations of Scales: What do They Tell Us, What do They Mean?* Manuscris.

Graziano, A. M., Raulin, M. L. (1993), *Research Methods. A Process of Inquiry*, Harper Collins College Publishers.

Gregory, R.J. (1992), *Psychological Testing. History, Principles, and Applications*, Allyn and Bacon, Massachusetts.

Greuter, M.A., Algera, J.A. (1989), *Criterion Development and Job Analysis*. în: Herriot, P. (ed.), *Assessment and Selection in Organizations. Method and Practice for Recruitment and Appraisal*, John Wiley & Sons, Chichester.

Grondlund, N. E. (1967), *Measurement and Evaluation in Teaching*, The MacMillan Company, New York.

Guilbert, J. J. (1987), *Ghid pedagogic pentru personalul sanitar*, Editura Medicală, Bucureşti.

Guilford, J. P. (1936), *Psychometric Methods*, McGraw-Hill Book Company,Inc., New York, London.

Guilford, J. P. (1965), *Fundamental Statistics in Psychology and Education*, McGraw-Hill Book Company.

Guion, R. M. (1974), *Open a New Window: Validities and Values in Psychological Measurement*, "American Psychologist", 29, pag. 287-296.

Guion, R. M. (1986), *Changing Views for Personnel Selection Research*, Lucrare prezentată la al 21-lea Congres Internaţional de Psihologie, Ierusalim, 15 iulie 1986.

Hall, C.S., Lindzey, G. ş.a. (1985), *Introduction to Thesis of Personality*, John Wiley & Sons, New York.

Hambleton, R.K., Swaminathan, H., Rogers, H.J. (1991), *Fundamentals of Item Response Theory*, Sage, Newbury Park.

Hammond, S. (1995), *Using Psychometric Tests*. În: Breakwell, G. M., Hammond, S., Fife-Schaw, C. (ed.), *Research Methods in Psychology*, Sage Publications.

Harris, W.G. (1987), *Computer-based Test Interpretations: Some Development and Application Issues*, "Applied Psychology: An International Review", 36 (3/4), pag. 237-247.

Hendriks, A. A. J. (1997), *The Construction of the Five-Factor Personality Inventory (FFPI)*, Teză de doctorat, Groningen.

Henry, W. E., Sanford, N. (ed.) (1972), *The California Psychological Inventory Handbook*, Jessey-Bass, Inc.

Hettema, J., Kenrick, D.T. (1992), *Models of Person-Situation Interactions*. În: Caprara, G.V., Van Heck, G.L. (ed.), *Modern Personality Psychology*, Harvester Wheatsheaf.

Hofstee, W. K. B., de Raad, B., Goldberg, L. R. (1992), *Integration of the Big Five and Circumplex Approaches to Trait Structure*, "Journal of Personality and Social Psychology", vol. 63, nr. 1, pag. 146-163.

Hofstee, W.K., Smid, N.G. (1986), *Psychometric Models for Analysis of Data from Personality Questionnaires*. În: Angleitner, A., Wiggins, J.S. (ed.), *Personality Assessment via Questionnaires*, Springer-Verlag, Berlin, Heidelberg.

Huba, G.J. (1987), *On Probabilistic Computer-based Test Interpretation and Other Expert Systems*, "Applied Psychology: An International Review", 36 (3/4), pag. 357-373.

Huberty, C. J. (1984), *Issues in the Use and Interpretation of Discriminant Analysis*, "Psychological Bulletin", vol. 95, nr. 1, pag. 156-171.

Jäger, R. S. (1993), *Measuring Reactions of Examiner and Examinee to Each Other and to the Psychodiagnostic Situation*. În: Nevo, B., Jäger, R. S. (ed.), *Educational and Psychological Testing: The Test Taker's Outlook*, Hogrefe & Huber Publishers.

John, O.P. (1990), *The "Big Five" Factor Taxonomy Dimensions of Personality in the Natural Language and in Questionnaires*. În: Pervin, L. A. (ed.), *Handbook of Personality: Theory and Research*, The Guilford Press, New York.

Johnson, J.A. (1996), *Predicting Observers' Ratings of the Big Five from Self-Report Inventories*, Comunicare prezentată la a VIII-a Conferinţă Europeană asupra Personalităţii, Ghent.

Klausnitzer, J.E. (1992), *Les tests d'intelligence*, Marabout, Alleur.

Kline, P. (1993), *The Handbook of Psychological Testing*, Routledge, New York.

Lafon, R. (1973), *Vocabulaire de psychopédagogie et de psychiatrie de l'enfant*, Presses Universitaires de France, Paris.

Landau, E. (1979), *Psihologia creativităţii*, Editura Didactică şi Pedagogică, Bucureşti.

de Landsheere, G. (1975), *Evaluarea continuă a elevilor şi examenele. Manual de docimologie*, Editura Didactică şi Pedagogică, Bucureşti.

Landy, F. L. (1985), *Psychology of Work Behaviour*, The Dorsey Press Chicago, Illinois.

Landy, F. L. (1986), *Stamp Collecting Versus Science. Validation as Hypothesis Testing*, "American Psychologist", vol. 41, nr. 11, pag. 1183-1192.

Landy, F. J., Rastegary, H., Thayer, J. F., Colvin, C. (1991), *Time Urgency: The Construct and its Measurement*, "Journal of Applied Psychology", 76, 5, pag. 644-657.

Lay, C.H. (1997), *Explaining Low-Order Traits Through Higher-Order Factors: the Case of Trait Procrastination, Conscientiousness, and the Specificity Dilemma*, "European Journal of Personality", vol. 11, pag. 267-278.

Li, H., Rosenthal, R., Rubin, D.B. (1996), *Reliability of Measurement in Psychology: From Spearman-Brown to Maximal Reliability*, "Psychological Methods", vol. 1, nr. 1, pag. 98-107.

Lindeman, R. H. (1978), *Evaluarea în procesul de instruire*, În: Davitz, J. R., Ball, S.(ed.), *Psihologia procesului educaţional*, Editura Didactică şi Pedagogică, Bucureşti.

Lindvall, C. M. (1967), *Measuring Pupil Achievement and Aptitude*, Harcourt, Brace & World, Inc., New York, Chicago, San Francisco, Atlanta.

Lohse, H., Ludwig, R., Röhr, M. (1982), *Statistische Verfahren für Psychologen, Pädagogen und Soziologen*, Volk und Wissen Volkseigener Verlag Berlin.

Lord, F. M. (1980), *Applications of Item Response Theory to Practical Testing Problems*, Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey.

Lord, F. M., Novick, M. R. (1968), *Statistical Theories of Mental Test Scores*, Addison-Wesley Publishing Company.

MacCallum, R.C., Mar, C.M. (1995), *Distinguishing Between Moderator and Quadratic Effects in Multiple Regression*, "Psychological Bulletin", vol. 118, nr. 3, pag. 405-421.

Marcus, S. (1997), *Empatie şi personalitate*, Editura ATOS, Bucureşti.

McCrae, R., Costa Jr., P.T. (1995), *Trait Explanations in Personality Psychology*, "European Journal of Personality", vol. 9, pag. 231-252.

Megargee, E. J. (1972), *The California Psychological Inventory Handbook*, Jossey-Bass Inc., Publishers, San Francisco, Washington, London.

Messick, S. (1995), *Validity of Psychological Assessment*, "American Psychologist", vol. 50, nr. 9, pag. 741-749.

Meuris, G. (1985), *La méthode des tests... à la croisée des chemins*, "Bulletin of the International Test Comission", 21, pag. 5-12.

Michell, J. (1997), *Quantitative Science and the Definition of Measurement in Psychology*, "British Journal of Psychology", 88, pag. 355-383.

Minulescu, M. (1996), *Chestionarele de personalitate în evaluarea psihologică*, Garell Publishing House, Bucureşti.

Mischel, W. (1968), *Personality and Assessment*, John Wiley & Sons, New York.

Most, R. (1987), *Levels of Error in Computerised Psychological Inventories*, "Applied Psychology: An International Review", vol. 36, nr. 3/4, pag. 375-383.

Murphy, K.R., Davidshofer, C.O. (1987 şi 1991), *Psychological Testing. Principles and Applications*, Prentice Hall International, Inc.

Nevo, B. (1993), *Face Validity Revised 1*, În: Nevo, B., Jäger, R. S. (ed.), *Educational and Psychological Testing: The Test Taker's Outlook*, Hogrefe & Huber Publishers.

Onicescu, O. (1970), *Procedee de estimare comparativă a unor obiecte purtătoare de mai multe caracteristici*, "Revista de statistică", 4.

Oosterveld, P., Vorst, H. C. M. (1995), *Methods for the Construction of Questionnaire; A Comparative Study*, Lucrare prezentată la a III-a Conferinţă Europeană de Evaluare Psihologică, Trier, 28-30 august 1995.

Ozer, D. J., Reise, S. P. (1994), *Personality Assessment*, "Annu. Rev. Psychol.", 45, pag. 357-388.

Paunonen, S.V., Jackson, D.N. (1985), *Idiographic Measurement Strategies for Personality and Prediction: Some Unredeement Promissory Notes*, "Psychological Review", vol. 92, nr. 4, pag. 486-511.

Păun, G. (1987), *Paradoxurile clasamentelor*, Editura Ştiinţifică şi Enciclopedică, Bucureşti.

Penner, L.A., Shiffman, S. ş.a. (1994), *Personality Processes and Individual Differences*, "Journal of Personality and Social Psychology", vol. 66, nr. 4, pag. 712-721.

Perugini, M., Gallucci, M. (1997), *A Hierarchical Faceted Model of the Big Five*, "European Journal of Personality", vol. 11, pag. 279-301.

Petrovsky, A. V., Yaroshevsy, M. G. (ed.) (1985), *A Concise Psychological Dictionary*, Progress Publishers, Moscow.

Piéron, H. (1952), *Vocabulaire de la psychologie*, P. U. F., Paris.

Piéron, H., Pichot, R. ş.a. (1952), *Méthodologie psychotechnique*, P.U.F.

Pitariu, H. (1978), *Psihologie industrială - Lecţii*, Universitatea Babeş-Bolyai, Cluj-Napoca.

Pitariu, H. (1985), *Îndrumător psihodiagnostic*, Universitatea Babeş-Bolyai, Cluj-Napoca.

Pitariu, H. (1994), *Managementul resurselor umane. Măsurarea performanţelor profesionale*, Editura ALL, Bucureşti.

Pitariu, H., Albu, M. (1993), *Inventarul Psihologic California: prezentare şi rezultate experimentale*, "Revista de psihologie", 3, pag. 249-263.

Pitariu, H., Albu, M. (1995), *Anxietatea faţă de calculator: cauze şi posibilităţi de reducere*, "Analele Universităţii Ecologice "Dimitrie Cantemir" din Târgu -Mureş", Secţiunea Psihologie, pag. 41-51.

Pitariu, H., Albu, M. (1996), *Psihologia personalului. Măsurarea şi interpretarea diferenţelor individuale*, Editura Presa Universitară Clujeană, Cluj-Napoca.

Pushton, J.P., Jackson, D.N., Paunonen, S.V. (1981), *Personality: Nomothetic or Idiographic? A Response to Kenrick and Stringfield*, "Psychological Review", vol. 88, nr. 6, pag. 582-589.

de Raad, B. (1996), *Personality Traits in Learning and Education*, "European Journal of Personality", vol. 10, pag. 185-200.

Reber, A. S. (1985), *The Penguin Dictionary of Psychology*, Penguin Books.

Richardson, M. W., Kuder, G. F. (1939), *The Calculation of Test Reliability of Rational Equivalence*, "The Journal of Educational Psychology", pag. 681-687.

Roşca, M. (1972), *Metode de psihodiagnostic*, Editura Didactică şi Pedagogică, Bucureşti.

Saucier, G., Goldberg, L. R. (1996), *Evidence for the Big Five in Analysis of Familiar English Personality Adjectives*, "European Journal of Personality", vol. 10, pag. 61-77.

Schmitt, M., Borkenau, P. (1992), *The Consistency of Personality*. În: Caprara, G.V., Van Heck, G.L. (ed.), *Modern Personality Psychology*, Harvester Wheatsheaf.

Schoonman, W. (1989), *An Applied Study on Computerized Adaptive Testing*, Swets & Zeitlinger B.V., Amsterdam/Lisse.

Schuerger, J.M., Zarrella, K.L., Hotz, A.S. (1989), *Factors That Influence the Temporal Stability of Personality by Questionnaire*, "Journal of Personality and Social Psychology", vol. 56, nr. 5, pag. 777-783.

Shackleton, V., Fletcher, C. (1984), *Individual Differences. Theories and Applications*, Methuen.

Shrock, Sh. A., Coscarelli, W. C. C. (1989), *Criterion-Referenced Test Development. Technical and Legal Guidelines for Corporate Training*, Addison-Wesley Publ. Comp., Inc., NY.

Silva, F. (1993), *Psychometric Foundations and Behavioral Assessment*, Sage Publications, Inc.

Speier, C., Frese, M. (1997), *Generalized Self-Efficacy as a Mediator and Moderator Between Control and Complexity at Work and Personal Initiative: A Longitudinal Field Study in East Germany*, "Human Performance", 10 (2), pag. 171-192.

Spielberger, Ch. D. (1988), *State-Trait Anger Expression Inventory*, Research Edition, Psychological Assessment Resources, Inc.

Spielberger, Ch. D., Gorsuch, R.L., Lushene, R.E. (1970), *STAI Manual for the State-Trait Anxiety Inventory*, Consulting Psychologist Press, Inc.

Steinberg, L., Thissen, D. (1996), *Uses of Item Response Theory and the Testlet Concept in the Measurement of Psychopatology*, "Psychological Methods", vol. 1, nr. 1, pag. 81-97.

Steyer, R. (1989), *Models of Classical Psychometric Test Theory as Stochastic Measurement Models: Representation, Uniqueness, Meaningfulness, Identifiability, and Testability*, "Methodika", vol. III, pag. 25-60.

Steyer, R., Ferring, D., Schmitt, M.J. (1992), *States and Traits in Psychological Assessment*, "European Journal of Psychological Assessment", vol. 8, 2, pag. 79-98.

Steyer, R., Schmitt, M. J. (1992), *Models of Latent State-Trait Theory*, În: Steyer, R., Gräser, H., Widaman, K. F. (ed.), *Consistency and Specificity: Latent State-Trait Models in Differential Psychology*, Preprint.

Stocking, M. L. (1987), *Two Simulated Feasibility Studies in Computerised Adaptive Testing*, "Applied Psychology: An International Review", 36, 3/4, pag. 263-277.

Swaminathan, H., Hambleton, R. K., Algina, J. (1974), *Reability of Criterion-Referenced Tests: A Decision-Theoretic Formulation*, "Journal of Educational Measurement", vol. 11, 4, pag. 263-267.

Tabachnick, B.G., Fidell, L.S. (1989), *Using Multivariate Statistics*, Harper Collins Publishers, New York.

Thorndike, R. L., Haagen, E. (1961), *Measurement and Evaluation in Psychology and Education*, John Wiley & Sons, Inc.

Traub, R. E. (1994), *Reliability for the Social Sciences. Theory and Applications*, vol. 3, Sage Publications.

Traxler, A. E. (1955), *Administering and Scoring the Objective Test*, În: Lindquist, E. F.(ed.), *Educational Measurement*, American Council on Education, Washington,D.C.

Truxillo, D. M., Donahue, L. M., Sulzer, J. L. (1996), *Setting Cutoff Scores for Personnel Selection Tests: Issues, Illustrations, and Recommendations*, "Human Performance", 9(3), pag. 275-295.

Tyler, L. E. (1963), *Tests and Measurement*, Prentice Hall, Inc.

Van de Vijver, F., Hambleton, R.K. (1996), *Translating Tests: Some Practical Guidelines*, "European Psychologist", vol. 1, nr. 2, pag. 89-99.

Van Der Maesen de Sombreff, P., Hofstee, W. K. B. (1989), *Personality Questionnaires and Inventories*. În: Herriot, P. (ed.), *Assessment and Selection in Organizations. Methods and Practice for Recruitment and Appraisal*, John Wiley & Sons, Chichester.

Vernon, P.E. (1965), *The Measurement of Abilities*, University of London Press.

Waller, N.G., Reise, S.P. (1989), *Computerized Adaptive Personality Assessment: An Illustration with the Absorption Scale*, "Journal of Personality and Social Psychology", 6, vol. 57, pag. 1051-1058.

Warr, P. (1997), *The Varying Validity of Personality Scales*, "Selection & Development Review", vol. 13, nr. 4, pag. 3-7.

Weiss, D. J., Vale, C. D. (1987), *Adaptive Testing*, "Applied Psychology: An International Review", 36, 3/4, pag. 249-262.

Zlate, M. (1994), *Big-Five - o tendinţă accentuată în cercetarea personalităţii*, Revista de Psihologie, 4, p. 305-315.

*** *Grand dictionnaire de la psychologie* (1994), Larousse, Paris.

*** *Le petit Larousse. Dictionnaire encyclopédique* (1993), Larousse, Paris.

*** *Standards for Educational & Psychological Tests* (1974 şi 1985), Publicate de American Psychological Association, Inc.

Tiparul la
Imprimeria «ATLAS–CLUSIUM»
Cluj-Napoca



ISBN: 973–555–200-0

Lucrarea de față are ca principal obiectiv familiarizarea cititorului cu metodele matematice și statistice care se folosesc la construirea testelor psihologice și la interpretarea scorurilor testelor. Alături de procedeele de calcul uzuale sunt prezentate metode noi, bazate pe teoria aproximării și pe teoria clasificării. Acestea îmbină ideile teoriei clasice a testelor cu cele ale teoriei răspunsului la itemi.

Căutându-se răspuns la întrebarea "Ce înseamnă un test bun?" s-a realizat o sinteză a ideilor cuprinse în lucrări clasice și moderne din domeniul teoriei testelor psihologice. Este propus și un punct de vedere nou asupra teoriei testelor, bazat pe teoria clasificării: testul psihologic este privit ca un instrument de ierarhizare și de clasare a subiecților, iar construirea unui test psihologic este asimilată operației de creare a unui clasament prin agregarea clasamentelor produse de itemi.